Digital Painting Analysis

Authentication and Artistic Style from Digital Reproductions

Jacobsen, Robert

Publication date:
2012

Document Version
Early version, also known as pre-print

AALBORG UNIVERSITY

Digital Painting Analysis

Authentication and Artistic Style from Digital Reproductions

C. Robert Jacobsen

Department of Mathematical Sciences
Aalborg University, Fredrik Bajers Vej 7G, 9220 Aalborg East

Preface

The present thesis is the outcome of my PhD study at Department of Mathematical Sciences, Aalborg University, Denmark, in the period from August 1st 2008 until October 31th 2011.

The main topic of the thesis is to present a method developed during my PhD studies for authenticating drawings and oil paintings based on digital reproductions. The method is capable of distinguishing authentic paintings of different artists from attempted imitations of the artist's painting style and is furthermore sufficiently stable to withstand realistic recording errors.

Furthermore, a study of more general statistical properties related to quantifying artistic style in included, as well as a suggestion for a metric between EMG signals.

Acknowledgements

I wish to thank my supervisor, Professor Morten Nielsen, for a good collaboration during the last three years; I have much enjoyed your ability to point out my highlights when I have been discouraged as well as your constructive feedback.

This project could not have started without the initialization and contacts of Professor Arne Jensen and would not have evolved as it has without his assistance.

I would like to thank Professor Daniel Rockmore for hosting me at Dartmouth College during the Fall of 2010. During my stay at Dartmouth I also had the pleasure of meeting and working with James Hughes and Daniel Graham. My thanks goes to all of you for a very inspiring collaboration and a hands on illustration of how to ask (the right?) questions.

In addition I would like to thank Oticon Fonden and Knud Højgaards Fond for financial support during my stay at Dartmouth College.

My gratitude goes also to friends and family – it is difficult to experience enlightening conversions and focusing off math on my own.

Finally, I wish to thank Mette for her support, love and patience – without you I would probably still be running in circles.

Aalborg, October 2011

C. Robert Jacobsen

In this revised version the thesis has been updated according to the initial review by the committee.

Aalborg, March 2012

C. Robert Jacobsen

Summary

This PhD thesis deals with mathematical and statistical modelling of visual art – an approach at quantifying artistic style and visual appearance.

The major part of this thesis treats with the problem of determining the authenticity of a painting or drawing from digital reproductions, that is, whether the artwork is indeed painted by the claimed artist.

In Chapter 2 an approach for authenticating oil paintings is introduced and tested on paintings related to the Danish painter Asger Jorn. The authentication method is based on hidden Markov modelling of second generation multiresolution representations of digital reproductions. We are able to correctly classify 11 out of the 15 photographed paintings.

Chapter 3 presents a thorough investigation of the effect of imperfections in the digital reproductions caused by various effects in the data acquisition. The experiments are performed with the authentication method from Chapter 2 and digital reproductions from the Museum of Modern Art in New York related to the Flemish artist Pieter Bruegel the Elder and (some of) his imitators. With this data set we can correctly classify 12 out of 13 drawings and since we classify both authentic and non-authentic drawings this is state of the art.

Chapter 4 includes a more general study of categorizing larger amounts of visual art by artists as well as visual similarity using higher order statistics.

Chapter 5 contains unpublished material in two parts: The first part is a suggestion for a metric between discrete EMG signals that overcomes limitations of usual approaches.

The second part uses the method from Chapter 2 to test the authenticity of drawings related to Pieter Bruegel the Elder from the National Gallery of Denmark. The authentication correctly classifies a drawing that is made by Pieter Bruegel and one that is made by his son Jan Bruegel, whereas the questionable works are distributed in between.

Appendix A is an introduction to the contourlet transform that is used in the authentication method.

Danish summary (Dansk resumé)

Denne ph.d.-afhandling omhandler matematisk og statistisk modellering af malerier og tegninger – en tilgang til at kvantificere kunstnerisk stil og fremtoning.

Hovedparten af afhandlingen vedrører problemet med at afgøre ægtheden af et maleri eller en tegning fra digitale billeder, det vil sige bestemme hvorvidt maleriet faktisk er lavet af den påståede kunstner.

I kapitel 2 introduceres en metode til bestemmelse af ægtheden af oliemalerier, der testes på malerier relateret til den danske kunstner Asger Jorn. I korte træk er metoden baseret på at benytte skjulte Markov-modeller til at modellere anden generations multiopløsningsrepræsentationer af digitale billeder.

Kapitel 3 er dedikeret til en grundig undersøgelse af effekten af skavanker i de digitale billeder som følge af dataindsamlingen. Eksperimenterne udføres med metoden fra kapitel 2 og digitale billeder fra Museum of Modern Art i New York af tegninger, der er relateret til den flamske kunstner Pieter Brueghel den Ældre og (nogle af) hans afslørede efterlignere. Med dette datasæt kan vi klassificere 12 ud af 13 tegninger og eftersom vi kan klassificere både ægte og uægte tegninger, er vores resultater state of the art.

Kapitel 4 inkluderer en mere generel undersøgelse, hvor vi kategoriserer store mængder visuel kunst i henhold til både ophavsmand og visuelle ligheder ved hjælp af højere ordens observatorer.

Kapitel 5 indeholder upubliseret materiale i to dele: Den første del er et forslag til en metrik mellem diskrete EMG-signaler, der ikke lider under begrænsningerne fra sædvanlige metoder.

Den anden del bruger metoden fra kapitel Chapter 2 til at teste ægtheden af tegninger relateret til Peter Brueghel den Ældre fra Statens Museum for Kunst. Metoden klassificerer tegningen af Pieter Brueghel korrekt som ægte, tegningerne af hans søn Jan Brueghel som falske og resten af tegningerne er fordelt derimellem.

Contents

CHAPTER 1

Introduction

Classification of paintings relative to their creator is an important part of its evaluation, as it may be change our perception of the painting artistically, culturally and monetarily.

A peculiar story to illustrate this point was in the news in 2009 (see e.g. [**25**]): In 2007 a gallery in New York sold a painting for 19,000 USD that at the time was thought to be a 19th century German painting. Then, two years later, art experts believed that the painting was made by Leonardo da Vinci and suddenly the estimated value had risen to an astounding 150 million USD.

The prize level of paintings by famous artists is of course also attractive for skilled forgers; if you succeed in selling your own work as e.g. a van Gogh painting you will have earned enough money for a lifetime.

A deception is truly successful if it goes unnoticed and no forger is interested in being detected, so the true extent of fake paintings is not known. However, the former director of the Museum of Modern Art in New York, Thomas Hoving, estimates that up to 40% of the paintings he has encountered through his career are forgeries [**34**]. Even if this prediction is pessimistic, the amount of money involved in dealing with art is of a magnitude that justifies the interest in exposing forgeries.

Determining the authenticity of paintings have traditionally been carried out by connoisseurs and art experts using a variety of methods. This includes determination of the painting's materials (e.g. paper type, watermarks, fibre composition), provenance, artistic content and analysis of color layers and drawing materials[1].

Sometimes it is possible to make absolute classification when determining the status

[1] Personal communication with Niels Borring, National Gallery of Denmark, and Teresa Østergaard Pedersen, Museum Jorn.

of a painting. This was for instance the case in the summer of 2011, when the National Gallery of Denmark revealed that a painting formerly attributed to Pieter Bruegel the Elder was painted after his death [**55**]. Such classifications are not always possible since forgers can be contemporaries of the artist they copy. Thus in many cases the classifications are based on methods that are more subtle.

As an attempt to detect forgeries it is of interest to be able to distinguish between the visual style of different artists. In fact, the connoisseur Bernard Berenson (1865–1959) insisted on the supremacy of visual evidence when authenticating paintings [**6**] and for many artists it is believed that the brushstrokes in a painting/lines in a drawing are characteristic.[2]

In a painting with distinctive brushstrokes, these form small ripples in a variety of orientations. The finish of a brushstroke can for some artists swing the balance in connection with their style, as mentioned in [**51**] where art experts at the Metropolitan Museum of Art in New York identified fake drawings formerly attributed to Pieter Bruegel the Elder by noticing that the end of the brushstrokes in certain drawings deviated too much from his usual style.

The validity of this assumption depends on the kind of painting in question; paintings in the e.g. photorealism style [**16**] do not have visible brushstrokes. In the other end of the spectrum is an artist like Vincent van Gogh who is known for his swirling brushstrokes [**14**]; indeed the experiments with digital authentication that involves van Gogh's paintings rely on the uniqueness of his brushstrokes [**1**, **3**, **4**, **42**, **46**]. Although it might not be as distinct as in van Gogh's paintings, brushstrokes can also be characteristic for other artists[3]. Of course we cannot prove that the assumption is correct, but our results confirm that it is reasonable.

In any case, when the tasks of authenticating paintings is performed by art experts the classification process is a costly and subjective procedure. Furthermore, if only a small number of art experts decide whether or not a painting should be considered authentic, this subjective truth is not timeless, but might be subject to change.

It is therefore of interest to automate part of this authentication process. Arguments that support this are improvements of the drawbacks of manual authentication, namely a reproducible and unbiased decision.

Although appealing, it is not realistic to construct a perfect automatic authentication procedure; just as when an art expert performs an evaluation, such a classification is associate with a degree of uncertainty. Quantifying our (un)certainty in a decision is the very heart of statistics, and an automatic authentication procedure must therefore be based on statistical methods.

In order to employ statistical decision theory for artist authentication it is necessary to find a suitable quantization of distinctive features. In this regard it is appealing to exploit the theory of uniqueness of brushstrokes; if we can quantify the brushstrokes of an artist mathematically and if the brushstrokes are indeed sufficiently unique for a given artist, this could serve as a basis for a classifier.

[2] To avoid the cumbersome paintings/drawings, the word "painting" will in the following be used as a synonym for both paintings and drawings, with brushstrokes meaning lines for drawings.

[3] Personal communication with Niels Borring, National Gallery of Denmark.

As a side note, it should be mentioned that also the decision of automatic classification methods might change if it turns out the ground truth used to make the decision is wrong, but this will be case no matter how a classification is performed.

A likely work flow involving an automatic authentication procedure could for instance be that the automatic procedure performs a presort of paintings; those that are obviously deemed forgeries need not be considered further. The paintings whose authenticity is more doubtful can be passed on to art experts who can then perform a finer inspection, eventually leading to a decision.

The majority of my focus during my PhD studies have been to develop a method for automatic authentication of paintings from digital reproductions and the rest of this introductory chapter is divided into three parts to introduce the work in this thesis: In Section 1.1 I will go through the chronological development of my PhD project; Section 1.2 is a more formal introduction including the mathematical setup and Section 1.3 is an outline of the rest of the thesis.

1.1 Development of project

The basis for the current work started long before this PhD project commenced, namely with the Portinari Project [**59**].

Candido Portinari (1903 - 1962) [**14**, **15**] was one of the most important cultural personalities in Brazil in the 20th century and amongst his activities he was a painter. Portinari is perhaps not well known outside of South America, but an illustration of his contemporary status is the fact that he has made the painting War and Peace in the delegates hall in the United Nations building in New York.

Portinari used lead based paint for his paintings and even though he had been warned by doctors that the paint was killing him, he continued his working style; this turned out to be, indeed, a lethal cocktail and he eventually died from lead poisoning at the age of 59 [**33**].

The paintings made by Portinari were not well documented at the time of his death; Portinari gave many of his paintings away, he did not keep records of all the paintings he sold and often he did not sign the date of his paintings [**33**]. This situation can be very beneficial for forgers: Is this newly discovered painting with Portinari's signature an authentic painting that has been in storage until now?

In 1979 Portinari's son, João Candido Portinari, started The Portinari Project to document the life of his father. A large part of The Portinari Project is to classify the images contributed to Portinari [**32**], to determine which images are in fact made by Portinari and which are forgeries.

João Candido Portinari is a mathematician by training and he was familiar with the theory that the brushstrokes of an artist is unique; it was therefore an appealing thought to quantify his father's brushstrokes and use these for classification.

In an oil painting the brushstrokes appear as ripples and this carries over to digital reproductions of the paintings. For those familiar with the Fourier transform, separating different ripples sounds like an easy task, since different ripples have

different frequencies.

The efforts in the Portinari Project involving the Fourier transform is mentioned in [**40**], but the experiments were not sufficiently successful for the classification task.

The Fourier transform has been around for a long time and is well suited for variety of signal processing tasks, but capturing the local information of brushstrokes seems to be outside of its reach.

Modelling small, localized ripples is on the other hand a job that the newer *wavelets* were designed to deal with and almost a decade after the experiments with Fourier analysis, a postdoc affiliated with the Portinari Project performed experiments with wavelets for authentication. These experiments came to an end with the unfortunate death of the postdoc, but the results are published in [**58**].

At some point João Candido Portinari got acquainted with Arne Jensen (now at Aalborg University) who then learned of the Portinari Project and their interest in an automatic procedure for authentication.

Later this also became my lead in to the field: In the end of 2008 I was in the situation of having to choose a new path for my PhD studies that would suit my professional background and interests in mathematical analysis and theoretical statistics/probability theory. Arne Jensen and my supervisor Morten Nielsen (also from Aalborg University) proposed this project that as an outlook seemed to offer a combination of my interests.

Arne and Morten are very knowledgeable in the field of harmonic analysis that includes the subject of the Fourier transform as well as wavelets and their generalizations. Along with the project proposal Arne and Morten also had an idea for a starting point, namely the use of generalizations of wavelets for extracting the brushstrokes in paintings.

The reason for involving more general methods than wavelets is that wavelets are originally developed for analysing ripples in one-dimensional signals and is extended to image analysis by analysing the horizontal and vertical ripples independently. Unfortunately the majority of realistic scenes does not fit into this rectangular framework and the need for retaining the good properties of wavelets at a larger variety of orientations have later spawned other transforms. The backbone of the work I have been doing on authentication have been based on using the so-called *contourlet* transform, which is one of the above mentioned extensions of the wavelet transform.

To my knowledge, we are the only team working on digital artist authentication that have utilised the contourlet transform; the advantage of using this transform will be explained in the later chapters.

On the other hand wavelets have appealed to many others working on artist authentication and on different artists, as can be seen in [**37**, **41**, **45**, **53**]. The way to utilize the wavelet transform also varies between the different approaches. Wavelets are not the only only tool used for digital authentication or more general tasks in visual stylometry; see for instance the overview articles [**29**, **56**].

As one can see from the publications mentioned, the field of digital artist authen-

tication is in its early stages. One reason for this is the sparse data available for testing authentication methods: Many museums have digital reproductions of their paintings[4] but it is not profitable to share digital reproductions of good quality with the public, as the sale of painting related merchandise (e.g. posters) is a significant source of income.

There is also the issue of having a sufficiently diverse data set. When training a statistical classifier to be able to distinguish between authentic images and forgeries it is rarely enough to have only authentic images – in that case we do not know how much a test image can deviate from the training data before it should be categorized as being *too* different to be authentic. For museums it is usually not of interest to acquire or keep paintings that are known forgeries; only in special cases will the forgeries be interesting in their own right.

Thus getting into the business of developing methods for authentication requires the right contacts to provide decent data. From his acquaintance with João Candido Portinari Arne Jensen had access to a large number of reproductions of Portinari's work, so this served as my initial testing data.

However, as my project evolved and I started testing my ideas, it became clear that the data provided from the Portinari Project was not of sufficiently homogeneous quality for reliable testing. To circumvent this problem we initially contacted Kunsten Museum of Modern Art in Aalborg who helped us acquire digital reproductions of paintings by the Danish painter Karl Larsen.

In order to have access to digital reproductions of paintings by an artist as well as paintings that are deliberately stylistically similar, we eventually contacted Museum Jorn in Silkeborg, Denmark, to expand our data set by photographing paintings of the Danish artist Asger Jorn and his collaborators (an short biography of Asger Jorn and his collaborators in the COBRA group can be found in [**6**, **13**, **15**]).

Though this was initially a set back, I think that it was beneficial in the long run for two reasons:

1. I learned a valuable lesson about quality checking of data.

2. We extended our focus from the pure authentication task to also consider how data quality and collection affects the authentication.

Regarding Item 1, I have later come to appreciate modern cameras not only due to the higher resolution (and quality) of the digital images, but also due to the extensive meta data they collect during recording. For instance it turns out that it is important to have the same number of pixels per real life area unit in the digital images; modern cameras estimate the distance to their focus points and thus it is possible to compensate for different number of pixels per area unit by scaling images according to their focus distances.

[4]Danish state-owned and state-subsidised museums are required to submit information about their collections to the central register Art Index Denmark, `https://www.kulturarv.dk/kid/`. Information about paintings are often accompanied by (low resolution) images.

I spent the fall of 2010 at the Department of Mathematics at Dartmouth College, NH, USA, to visit Professor Daniel Rockmore. I first became acquainted with Daniel Rockmore from two important contributions to digital artist authentication, namely [**35**, **45**].

In [**45**] the authors used wavelet transforms for authentication, by comparing suitable statistics computed from the wavelet coefficients.

The second article [**35**] utilizes a completely different idea for authentication based on *sparse coding* and received a lot of attention in the media, see e.g. [**51**, **54**].

The application of sparse coding to authentication is the appealing idea of computing a basis from authentic paintings of a given artist and then see how well this basis fits the test painting; a better fit means that it is more likely that the test painting is also made by this artist.

The two papers [**35**, **45**] co-authored by Daniel Rockmore both test the authentication methods on drawings by Pieter Bruegel the Elder (little is known about Bruegel's life, but short biographies and introductions to his artistic contributions can be found in [**14**, **36**]).

Luckily it turned out that the method I had developed was also capable of separating these Bruegel drawings; indeed the separation of the Bruegel images from the forgeries is clearer than the separation we have in our self obtained data set related to Asger Jorn.

While at Dartmouth I was introduced to a wealth of new perspectives on mathematics and art – not only restricted to authentication. This involves in particular modelling artistic style and exploring higher order features of images; this is elaborated in Section 1.3 where the main work of this thesis is presented.

After returning from the USA in the early 2011 my work has involved a few more twists influencing this thesis.

In the spring I attended a series of talks by people from the Department of Mechanical and Manufacturing Engineering at Aalborg University about modelling the human body. After the talk one of the speakers mentioned that they were looking for a method to validate some of their simulated models. The validity of the model basically boils down to judging whether or not two one-dimensional signals are sufficiently alike.

I was intrigued by this problem and I came up with an idea of how to do this – the approach will covered later in Section 5.1.

In the 11th hour of my project I read in a Danish newspaper article [**55**] that the National Gallery of Denmark had exposed a painting formerly attributed to Pieter Bruegel the Elder as a forgery. The exposure had been performed using traditional "hands-on" methods and eventually by determining that the materials in the painting were younger than Pieter Bruegel, as mentioned earlier.

Since I have had success with distinguishing between drawings of Bruegel and those of his imitators, I contacted the National Gallery to commence a collaboration. During the summer of 2011 I was provided with digital reproductions of paintings by various artists and artistic style; this was a opportunity of testing the real

application value of our authentication method – but it turned out that it also required extra work. The extra work arose from the difficulties of comparing two sets of digital reproductions that were completely different; the drawings recorded are naturally different, but also the environments and cameras used for digitization were different.

Towards the end of this project I believe I have been able to compensate for these differences and I have obtained good results classifying these data. The work in this regard is not over, however, since additional automation is desirable.

1.2 Methodology

This section contains a brief description of the mathematical context for the thesis.

As mentioned above, the assumption we have used in this work is that distinctive features of artistic style can be found in the high frequency part of painting or drawing. The basis for our analyses are digital reproductions of the paintings; in suitable units a digital image is a real function on $\mathbb{Z}^2$ that is non-zero on a finite, rectangular area.[5]

In this discrete, infinite setup the Fourier transform is an operator $\mathcal{F} : \ell^2(\mathbb{Z}^2) \to L^2([-\pi, \pi]^2)$ that maps the digital image into a function that is 2π-periodic in both variables.

In this setup, the low frequency part is the frequencies within a neighbourhood of the origin (where the suitable neighbourhood depends on the context) and the high frequency part is its complement in $[-\pi, \pi]^2$.

The classification used for authentication presented in Chapter 2 and 3 is based on two main ingredients; the contourlet transform for selecting the high frequency content and hidden Markov models used for statistical modelling of the contourlet transform. The subsequent analysis uses the fitted hidden Markov models for classification and this part is of course important as it determines the success of our method, but it is more a superstructure: The tools we have employed for the classification yields good results, but it is also possible that we could make better use of the information captured and modelled by the contourlet transform and hidden Markov models, respectively. Consequently, I will here give short introductions to these main tools; the contourlet transform is treated more in-depth in Appendix A and is also motivated for our experiments in Section 2.3.1; hidden Markov models are treated in Section 2.3.2.

Hidden Markov models are parametric models and as such, we perform parameter estimation when fitting them to our data. To perform maximum likelihood estimation for parameter estimation, we employ an Expectation Maximization (EM) algorithm, and I will introduce the general algorithm and the specific algorithm is elaborated in connection with the hidden Markov models in Chapter 2.

[5]The reason for considering this infinite dimensional setup is to avoid cluttering of the conceptual introduction with practicalities like boundary corrections. From a personal point of view I also find the Fourier transform easier to visualize in this setup.

The work in Chapter 4 has a foundation in the same sense as Chapter 2 and 3, which is *sparse coding*. Due to space constraints in the proceedings, our article in Chapter 4 did not introduce sparse coding, so I will also give a brief introduction to sparse coding here.

Contourlet transform Here it is assumed that the reader has some familiarity with multiresolution analysis. General introductions to multiresolution analysis can be found in many references, see e.g. [**31**] for a mathematical introduction and [**60**] for a signal processing perspective.

The contourlet transform is based on a multiresolution analysis of $L^2(\mathbb{R}^2)$ and is implemented with a double filter bank[6]; one for separating low and high frequencies and one for selecting directional frequencies.

After each application of the contourlet filter bank, we are left with two new images; an approximation in the form of a low frequency subband which we downsample and "the rest", i.e. a high frequency subband, which we refer to as details. The contourlet filter bank can the be applied successively to the approximation leaving a coarser approximation and a new level of details. We refer to the details at level l as the details we obtain after applying the contourlet transform l times.

The true justification for our interest in the contourlet transform is not its ability to perform frequency selection; this part of the transform builds on early work [**8**]. What makes the contourlet transform appealing is that we we can divide the high frequency subbands further into *directional* subbands, containing high frequencies in a multitude of directions.

The general idea behind a multiresolution analysis, and thereby the contourlet transform, is illustrated in Figure 1.1.

Hidden Markov model In order to use the contourlet transforms for classification, we need to model them in a sensible way. The choice of hidden Markov models is motivated in Section 2.3.2; here the main idea is described.

Referring to Figure 1.1, the starting point for the hidden Markov modelling of contourlet transforms is that the different levels of details are inter dependent; they capture the high frequency details, but at different resolutions. Significant edges are persistent over different resolution levels and therefore part of the high frequency content on these levels, meaning that we model the edges.

Due to the downsampling between each level, the individual details span fewer pixels on the higher levels. More precisely, a single pixel on level l represents four pixels on level $l-1$ and these are the relations we model with hidden Markov models.

The term "hidden" is used because it is convenient to associate a so-called hidden state with every pixel that we cannot observe – its distribution has to be inferred

[6] A two dimensional filter is an element of $\ell^2(\mathbb{Z}^2)$ with a bounded frequency support. The corresponding filter bank is an operator on $\ell^2(\mathbb{Z}^2)$ that implements the filter through point wise multiplication in the frequency domain or convolution in the spatial domain. The output(s) from a filter bank are called subbands, which is a certain selection of frequencies from the image.

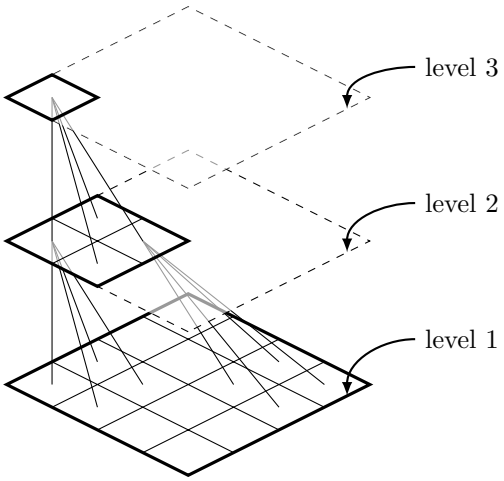


Figure 1.1: Illustration of the idea in a multiresolution analysis – which is also the basis for the contourlet transform – that provides us with details of an image at different levels. The lower levels have higher resolution and contains higher frequency content. The lines connecting the levels show the pixels that are related and modelled by hidden Markov models.

from the (observed) contourlet transform itself.

Expectation Maximization (EM) algorithm The EM algorithm is general tool for performing maximum likelihood estimation in a model with hidden variables and is described in many books, see e.g. [**5**, **24**, **30**]. Here we follow the notation of [**24**].

In statistical terminology our sample consists of N contourlet coefficients $\mathcal{D}_o = \{c_i\}_{i=1}^N$. In our observation model we augment each of the observed coefficients with a hidden state s_i indexed in the same manner; the hidden states are collective written as $\mathcal{D}_u$.

Letting $\boldsymbol{\theta}$ denote the parameters of the hidden Markov model, the complete log-likelihood is

$$p(\mathcal{D}_o, \mathcal{D}_u; \boldsymbol{\theta}) \tag{1.1}$$

whereas the observed log-likelihood is

$$p(\mathcal{D}_o; \boldsymbol{\theta}) = \int p(\mathcal{D}_o, s; \boldsymbol{\theta}) ds. \tag{1.2}$$

The likelihood (1.2) is often difficult to maximize, whereas (1.1) can be easier, as is the case for the hidden Markov models we consider. However, to work with (1.1) we need the observations $\mathcal{D}_u$. The EM algorithm circumvents this problem by estimating $\mathcal{D}_u$ from $\mathcal{D}_o$ and the observation model with a chosen set of parameters.

More formally, for a fixed set of parameters $\boldsymbol{\theta}^i$ we introduce the function

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^i) = \mathrm{E}\big(\ln p(\mathcal{D}_o, \mathcal{D}_u; \boldsymbol{\theta}) | \mathcal{D}_o; \boldsymbol{\theta}^i\big), \tag{1.3}$$

where the expected value is with respect to $\mathcal{D}_u$. The function Q is thus the complete likelihood where the unobserved values are replaced with their expected values from the observations $\mathcal{D}_o$ and the current set of parameters $\boldsymbol{\theta}^i$.

With this notation the general EM algorithm is as follows.

Algorithm 1: Expectation-Maximization algorithm

Input: Initial parameter estimate $\boldsymbol{\theta}^0$, tolerance T
Output: Maximum likelihood estimate $\widehat{\boldsymbol{\theta}}$
Initialization: $i = 0$
repeat
  $i \leftarrow i + 1$
  E step: compute $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{i-1})$ from (1.3)
  M step: $\boldsymbol{\theta}^i \leftarrow \operatorname{argmax}_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{i-1})$
until $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^i) - Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{i-1}) \leq T$
return $\widehat{\boldsymbol{\theta}} \leftarrow \boldsymbol{\theta}^i$

What makes the EM algorithm valuable is that after a full sweep we are guaranteed that the likelihood has not decreased, i.e.,

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^i) \geq Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{i-1}).$$

This implies that the sequence $\{\boldsymbol{\theta}^i\}_{i \geq 0}$ converges to a local maximum likelihood estimate or a saddlepoint of the likelihood function (1.2) [**5**].

Sparse coding A major part of the visual processing system in human brains is the visual cortex which is divided into different areas, named V1 through V5. The first part, V1, is also called the primary visual cortex and – as the name suggests – this part accounts for a major part of the neural processing.

There is reason to believe that the human brain decomposes natural scenes into basic building blocks (i.e., a basis, or more generally a dictionary) in such a way that any given patch of the scene can be represented with as few building basis functions as possible, i.e. as sparsely as possible (see e.g. [**50**]). Sparse coding was introduced in [**49**, **50**] as a model for V1 that incorporates this feature. As it has later turned out, this criteria of sparseness is very useful in a wide variety of applications.

Let us now consider this in a mathematical framework. As above we consider an image to be a real function on $\mathbb{Z}^2$. The goal in sparse coding is to determine a set of functions that locally span I as sparsely as possible. This means that we consider patches of I, $p : \{1, \ldots, n\}^2 \to \mathbb{R}$,

$$p(x, y) = I(x - k_1, y - k_2)$$

for some $k_1, k_2 \in \mathbb{Z}$. Let $\{\phi_i\}_{i=1}^N$ denote the set of functions $\phi_i : \{1, \dots, n\}^2 \to \mathbb{R}$ that we want to determine. In mathematical formulation, the functions must span the patches sparsely, i.e., there exists $a_1, \dots, a_N$ such that

$$p = \sum_{i=1}^{N} a_i \phi_i \tag{1.4}$$

with the vast majority of a_i's having a negligible numerical value for the majority of patches from I. The functions $\{\phi_i\}_{i=1}^N$ are not necessarily linearly independent.

For computational purposes these demands are formulated as follows: For a given patch p and functions $\{\phi_i\}_{i=1}^N$, the construction error of a sequence of weights $\{a_i\}_{i=1}^N$ is

$$\sum_{x,y=1}^{n} \big(p(x,y) - \sum_{i=1}^{N} a_i \phi_i(x,y)\big)^2 \tag{1.5}$$

and the sparsity is proportional to

$$\sum_{i=1}^{N} S(a_i), \tag{1.6}$$

where S is a penalty function; often $S(x) = \log(1 + x^2)$, see Figure 1.2. The reason why this particular penalizing function is well suited for the task is that it heavily penalizes non-zero values, but its (uneven) derivative quickly reaches its global maximum/minimum. This means that if a particular ϕ_i in (1.4) is significant, the penalization of a_i does not increase out of proportion once a_i reaches a certain level.

The ϕ_i's are found by simultaneously minimizing (1.5) and (1.6); an example of what such basis functions typically look like is presented in Figure 1.3.

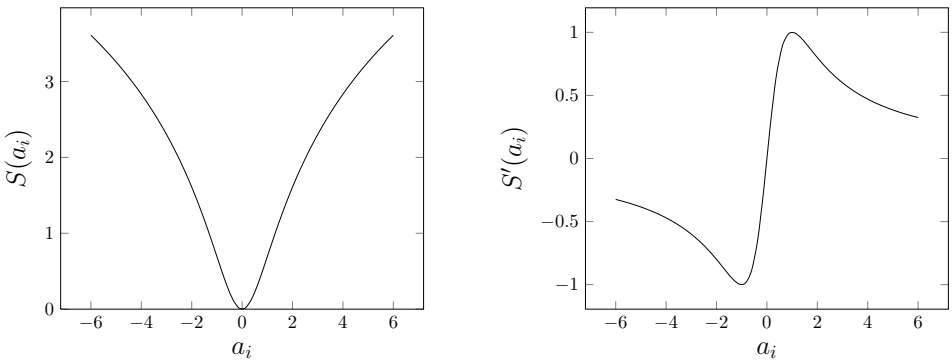


Figure 1.2: The penalty function S we use in (1.6) and its derivative.

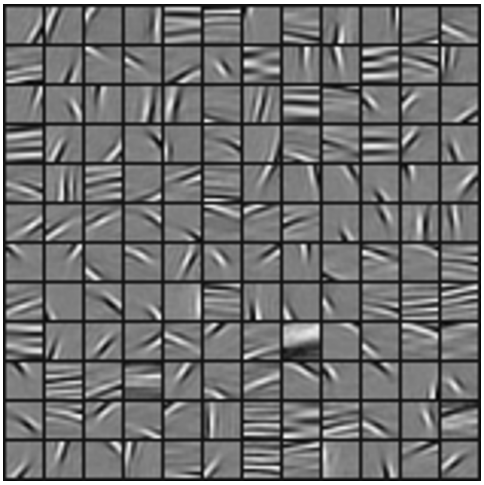


Figure 1.3: Example of a sparse coding basis. These functions are trained from patches from the drawing by Pieter Bruegel the Elder seen in Figure 3.1. Each of the functions are normalized to take values in the interval $[-1, 1]$, where -1 is the darkest and 1 is the lightest.

1.3 Outline

Chapter 2 to 4 presents the papers that constitute the main part of this thesis; an outline of each of the papers is presented below and the publication details of the individual papers is presented in the beginning of the corresponding chapter.

The papers are presented here in (extended) journal form – including a separate bibliography. This means that repetitions occur in the later chapters, but also that they can be read independently.

Chapter 2 presents the method I have developed for authentication of paintings from digital reproductions. The method models contourlet transforms with hidden Markov models and uses the parameters of these models for classification using support vector machines.

We test the method on paintings by the Danish artist Asger Jorn as well as publicly available paintings by Charlotte Caspers. Furthermore we investigate the effect of different means of digitization.

This is an extended version of the submitted paper with more technical elaborations as well as reflections on the visual characteristics of the paintings in our analysis.

Chapter 3 is a study of the robustness of the authentication method presented in Chapter 2. The effect of digital image quality and means of acquisition is a largely unexplored area in digital authentication. We investigate how factors related to acquisition of the digital reproductions affect the conclusions of our method.

This chapter is extended with technical details compared with the submitted paper.

Chapter 4 is a paper presented at the European Signal Processing Conference 2011 (EUSIPCO), where we investigate the use of higher order statistical features for different ways categorizing visual art. Authentication aims at categorizing art according to the painter, but here we also look into categorization according to visual perception.

Chapter 5 contains two sections with work carried out near the end of my PhD studies that I have not finished. The first section presents a suggestion for metric between samples of discrete signals. The application of this metric is illustrated on data collected from EMG measurements.

The second section is an analysis of paintings from the National Gallery of Denmark, where we determine their authenticity with the methods presented in Chapter 2.

Appendix A is an introduction to the contourlet transform that has been used extensively in my work.

CHAPTER 2

Authentication of paintings using hidden Markov modelling of contourlet transforms

Publication details

Co-authors: Morten Nielsen*

* *Department of Mathematical Sciences*
Aalborg University

Journal: Submitted to EURASIP Signal Processing, special issue on Image Processing for Art Investigation.

This is an extended version of the submitted paper with more artistic considerations and elaboration of technical details.

Abstract:
Authentication of paintings is the task of determining if a painting is indeed made by the claimed artist. In this paper we present a method for digital authentication of paintings, which is the task of performing authentication from digital reproductions of paintings.

Our method is framed around modelling contourlet transforms of the digital reproductions with hidden Markov models. Using the contourlet transform in this field of classification is a new approach motivated by the contourlets' efficiency in representing piecewise smooth contours such as brushstrokes.

To test our method we have used paintings related to the Danish painter Asger Jorn; all paintings are recorded in multiple digital images *and* by two different cameras. With multiple sources we are able to get insight into the robustness of our method against different means of acquisition.

Through a cross-validation of the data set by one of the cameras we are able to correctly classify 39 out of 44 images; based on this classifier we can correctly classify 30 out of 36 images in the other data set.

Keywords
Visual stylometry; authentication; contourlet transform; hidden Markov models; classification

2.1 Introduction

The aim of visual stylometry of paintings is to quantify the artistic style of a painter and thus get an insight into the development of his style throughout his career. In the present paper we study quantification for the purpose of authenticating paintings based on digital reproductions: Is it possible to determine the authenticity of a painting based on digital photos?

Traditionally the task of authenticating paintings has been performed by art experts and connoisseurs with a profound and detailed knowledge of the artist and his contemporaries. However, this method will always involve some subjectivity and therefore the experts do not always agree. So assisting the authentication by more unbiased automatic methods may – besides being a topic of mathematical interest – also be valuable for the art community.

All automatic authentication methods works on digital reproductions of the paintings and the fundamental idea is to extract a number of features from the digitized paintings that are sufficiently expressive to distinguish the styles of different artists. Previous attempts have been made to perform automatic authentication of paintings by training models to separate paintings that are known to be authentic (i.e., painted by the claimed painter) from known forgeries.

The authentication task has been tested most thoroughly on paintings by van Gogh [**3**, **4**, **5**, **28**, **30**, **33**], Pieter Bruegel the Elder [**25**, **26**, **31**] and Jackson Pollock [**1**, **27**, **37**, **38**, **39**], but also others have been considered [**29**, **35**].

Our work in this area is motivated by the theory that a forger can reveal himself

by having more hesitant brushstrokes [**28**]. An explanation for this behavior is that when a forger imitates another painters style, the motion is unnatural and stammering. With this assumption it seems attractive to explore the fine details in paintings for the authentication purpose.

Several of the above mentioned methods employ multiresolution analysis to extract relevant features from the paintings. Both wavelets [**28**, **29**, **31**, **33**] and curvelets [**25**] have been applied for this task. The reason for applying these signal processing tools is that they are (hopefully) able to detect subtle differences in the details between authentic paintings and forgeries at different scales in the paintings.

In this paper we work with the contourlet transform [**15**] of the digital photographs. The reason for this choice of multiresolution analysis is that the atoms of the contourlet transform resemble contours well, making it likely to also represent the brushstrokes well and thereby the (subtle) differences between paintings made by different artists.

The contourlet transform of the paintings is modelled by a hidden Markov model [**32**] and by exploiting the differences between the hidden Markov models of different artists, we are able to predict the affiliation of new images. Our method is a development of the Princeton approach in [**28**], where complex wavelets are modelled by hidden Markov models instead of contourlets.

The rest of the paper is organized as follows: In Section 2.2 we present the data used in the experiments. Section 2.3 describes the tools we utilize for our classification task. This includes the contourlet transform and how we model these with hidden Markov models to distinguish between paintings by different artists. In Section 3.4 we present the most important results of our experiments. Section 2.5 discusses the interpretation of the presented results and Section 2.6 contains our conclusions.

2.2 Artists used in our experiments

We have worked with paintings related to two different artists, as described in the subsections. The acquisition of the digital reproductions have varied – for the data we have gathered ourselves this is a deliberate choice since we want to test the robustness of our method and in order to avoid the problems mentioned in Section 2.2.2.

Each colour image with Red, Green and Blue (RGB) components was converted to grayscale (gray = 0.2989 R + 0.5870 G + 0.1140 B) as in [**25**, **31**] in an attempt to reduce our models dependency of color variations in the images.

Each image was furthermore divided into square patches with a side length of 1024 pixels. Since the side lengths of a given image was rarely an integer multiple of 1024, the patches were taken from the central area of the image – sometimes with a small overlap between the individual patches.

2.2.1 Asger Jorn

Our primary dataset is related to the Danish painter Asger Jorn (1914 - 1973) and consists of paintings by Asger Jorn and his collaborators/apprentices. The images were provided courtesy of Museum Jorn, Silkeborg, Denmark.

We photographed the paintings with two cameras with very different technical specifications: A Canon Powershot G2 and a Nikon D90 with an AF-S Nikkor 50mm f/1.4G lens. We used two different cameras to test the robustness of our classification procedure to the means of acquisition and digital format. The Canon camera recorded images in JPEG format and the Nikon camera in raw format which we then exported to a lossless TIFF image. We have not considered the effects of JPEG compression in the following work; the motivation for using two very different cameras was to investigate if this influenced our method, cf. the situation mentioned in Section 2.2.2. The influence of the specific compression scheme and its parameters was not our initial focus and due to the success reported in Section 2.4.1 we have postponed this investigation to later work.

The distance between the Canon camera and the paintings was consistent for all photos, however, this was not the case as to the Nikon camera. The photos taken by the Nikon camera were digitally corrected afterwards for the inconsistency in distance, i.e., for the inhomogeneous number of pixels per physical area – this was possible since the camera recorded the distance to its focus point

The paintings were photographed in their display positions on the walls in the museum where the lighting was homogeneous.

The paintings photographed are listed in Table 2.1.

Not all paintings could be captured in a single photograph of sufficiently high resolution per physical area unit, so these were photographed in parts – leaving us with a total of 44 digital images of the 15 paintings.

Figure 2.1 presents an example of a painting by Asger Jorn and a painting by one of his collaborators, Helmut Sturm.

2.2.2 Charlotte Caspers

The authors of [**28**] concluded that they could decide on the authenticity as to a series of van Gogh paintings of which some were known to be forgeries. However, as the results of [**28**] were found to relate more to the cameras used for acquisition, the authors continued their research [**33**], but this time they made sure that they had full control over the entire process of acquiring the data: They asked the conservator Charlotte Caspers to paint a series of small paintings and then to copy them herself, and afterwards the digital reproductions were obtained with the same scanner. The digitized versions of these paintings can be found online [**9**].

The observation made by the authors of [**28**, **33**] is important as an authentication method with limitations as to the acquisition and digital format is not very useful.

Another complication when quantifying the artistic styles of painters is their use of different types of paint and canvas: We have not yet investigated the influence

Artist	Title	Year	Catalog no.
Asger Jorn	Automolok	1948	1986/0001
Asger Jorn	Euphorisme	1970	1971/0274
Asger Jorn	Hoved	1935	1958/0043
Asger Jorn	Henning. Figurstudie	1933	1972/0205
Asger Jorn	Portræt. Bodil	1961	—
Asger Jorn	Grand ventre - incendie	1953	1958/0210
Asger Jorn	Digteren Jens August Schade	1937-44	2008/0001
Asger Jorn	Prete Alla Spiaggia	1957-59	1977/0001
Asger Jorn	Trolden og fuglene	1944	1962/0183
Asger Jorn	Le Vent Nous Emporte	1970	1988/0001
Asger Jorn	Uden Titel. 1946	1946	1961/0116
Asger Jorn	Uden Titel	1956-57	1961/0115
Jacqueline de Jong	Admiration de la reine de vert	1961	1961/0119
Asger Jorn/Enrico Baj	Uden titel	1958	1958/0290
Helmut Sturm	Uden titel	1961	1961/0114

Table 2.1: The paintings related to Asger Jorn used in our experiments. The first column is the name of the artist, the second column is the original name of the painting, the third column is the year(s) of production and the fourth column is the catalogue number of the paintings at Museum Jorn. Additional information can be found at Art Index Denmark, `https://www.kulturarv.dk/kid/`. The painting "Portræt. Bodil" is displayed at Museum Jorn, but belongs to a private collector and therefore it has no catalog number.

(a) Asger Jorn, *Hoved [Head]*, 1958, oil on canvas pasted on laminate, 35.5 cm × 26 cm. © Museum Jorn, Silkeborg.

(b) Helmut Sturm, *Uden Titel [No Title]*, 1961, oil on canvas, 75.5 cm × 80.5 cm. © Museum Jorn, Silkeborg.

Figure 2.1: Examples of the paintings used in our experiments. Helmut Sturm was a collaborator of Asger Jorn.

of materials thoroughly.

2.3 Methods

In the first part of this section we explain our choice of model and how we fit the models to our painting data. The last part of the section is concerned with how we utilize our models in detecting forgeries.

The training part of our authentication method works in basically three steps:

1. Make a contourlet transform of patches from the digital images, Section 2.3.1.

2. Model the contourlet coefficients by a hidden Markov model. Hidden Markov models are very rich models that have proved to be useful for classification tasks, see e.g. [**13**, **32**, **34**], Section 2.3.2.

3. Construct a classifier from the hidden Markov models, Section 2.3.3.

The contourlet transform and hidden Markov models are both well-known tools; we include brief introductions to make the paper more self contained and to justify our choices. Furthermore, we set the notation used in later sections.

The trained classifier is used to classify new images.

2.3.1 The contourlet transform

In this section we briefly introduce the ideas of the contourlet transform and our motivation for using it in the present work. The reader is referred to the original article [**15**] for a more thorough exposition.

The contourlet transform offers a multiresolution decomposition of digital images with a user specified number of directional highpass subbands.

The multiresolution decomposition is obtained by successively computing coarser versions of the given image by applying a low pass filter in the form of a Laplacian pyramid. The highpass part that is filtered out at each level of the multiresolution is divided into directional subbands, i.e., subbands that contains high frequency content in a limited band of directions.

The contourlet transform is constructed using a multiresolution analysis and therefore, after each application of the low pass filter, we downsample the obtained low pass image by a factor of two in each direction. This implies that at each level of the decomposition the number of pixels in the image is a quarter of that at the former level.

The contourlet transform has a number of attractive features that have served as a motivation for using it in the present work. The *individual* features are not unique to the contourlet transform, but the fact that they are all present is what makes the contourlet transform appealing.

Directionality As stated in the beginning of the section, the highpass subband from each level is divided into directional subbands. Here the point is that the *number* of directional subbands is specified by the user. Having this choice is different from e.g. wavelets, where the high frequency content is distributed between a fixed set of directional subbands (vertical, horizontal and diagonal in the case of ordinary wavelets).

The directionality can also be obtained by other transforms, e.g. by Gabor wavelets or curvelets [**7**].

Optimal representation An alternative view on the contourlet transform is that it provides a frame $\{\phi_i\}_{i\in J}$ that allows us to represent an image I as

$$I = \sum_{i\in J} c_i\phi_i. \tag{2.1}$$

When quantifying how well a transform performs at representing images, we measure how well I can be represented using only a subset of the functions in (2.1). More formally, we let

$$I_M = \sum_{i\in J_M} c_i\phi_i,$$

where $J_M \subseteq J$ is the set of indices of the M numerically largest coefficients from (2.1). If the representation (2.1) is efficient, I's behavior is captured by a few ϕ's and $I_M \approx I$, quantified by a suitable norm $\|I - I_M\|$.

In [**16**] it was established, that in a continuous setting and if I is sufficiently smooth, $\|I - I_M\|^2$ cannot decay faster than M^{-2}. The curvelet and contourlet transform almost obtain this bound [**8**, **15**], as

$$\|I - I_M\|^2 \leq \text{constant} \cdot M^{-2}(\log M)^3, \tag{2.2}$$

provided that the number of highpass directions double at every other scale. This requirement is known as parabolic scaling and implies that the ϕ_i's in (2.1) have elongated supports that make them excel at capturing the edges (or contours) in images.

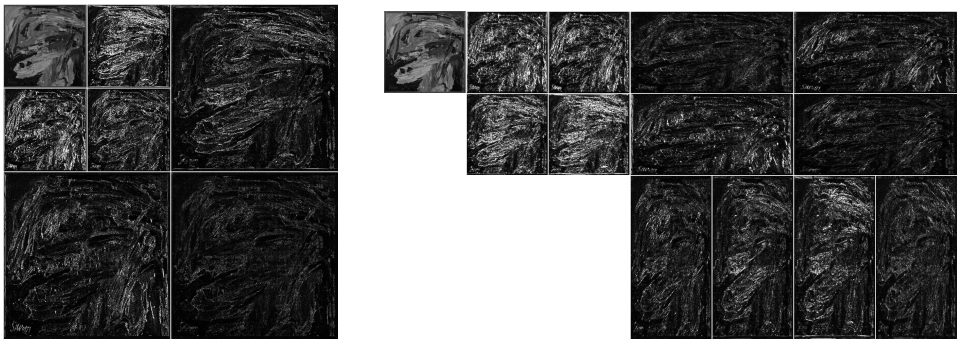
To the best of our knowledge, no transform has been found that outperforms the contourlet (and curvelet) transform in this respect.

Discrete construction The contourlet transform is *designed* to work on discrete data and hence the implementation for digital images is seamless.

This is the major difference between the curvelet and the contourlet transform; the curvelet transform is designed to work on continuous data and is adapted to discrete data via a sampling grid.

By using a directional multiresolution transform it is our hope that we get a good representation of the brushstrokes captured in a digital image at different resolutions.

An example of a possible contourlet transformation of an image is seen in Figure 2.2 – along with a wavelet transformation for comparison.



(a) A two level wavelet decomposition of image.

(b) A two level contourlet decomposition of image.

Figure 2.2: Comparison of wavelet and contourlet transformation by a two level decomposition of the grayscale version of the painting in Figure 2.1b. In the high-pass subbands of the contourlet transformation we capture details at a variety of orientations, as opposed to the three fixed orientations of the wavelet transform.

2.3.2 Hidden Markov modelling of contourlet transforms

Figure 2.3 illustrates the situation we want to model, i.e., how the coefficients evolve through the levels of the multiresolution decomposition. The coefficients that are connected in Figure 2.3 are at the same spatial locations in the image, but at different levels of resolution.

Referring to Figure 2.4, they share the same major direction, as the children levels together span the same direction as their parents level.

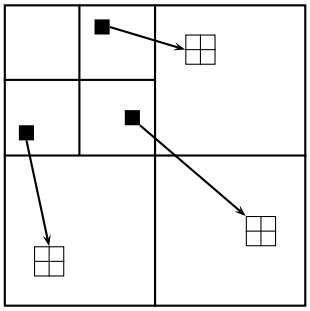
With these common characteristics it seems plausible that the coefficients are related; if the parent coefficient is large, indicating the presence of an edge, it is likely that (at least) one of the children coefficients is large as well – and vice versa. This relationship is, however, mostly local, as edges do not usually (dis)appear spontaneously at a fine level if they are (not) present at their parent level.

From (2.2) we have that the contourlet transform captures an image with few large coefficients, while the majority of coefficients are insignificantly small. So the distribution of coefficients at a particular subband is highly peaked around zero and has heavy tails.

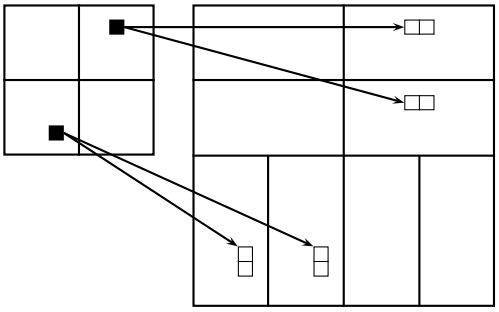
As it has been demonstrated in [**32**], such a distribution is well approximated by a *mixture* of normal distributions.

The hidden Markov model we utilize in our modelling was introduced for contourlets in [**32**] and captures the essential properties mentioned above in the following manner:

- An individual coefficient o_i is a realisation of a stochastic variable O_i whose distribution is a mixture of normal distributions. For each coefficient we also



(a) Coefficients in a two level wavelet transform of an image.



(b) Coefficients in a possible two level contourlet transform of an image.

Figure 2.3: Parent-children relationship for wavelets and contourlets. The black coefficients are parents of the white coefficients. The parents reside at the same same spatial location as their children, but at a coarser resolution.

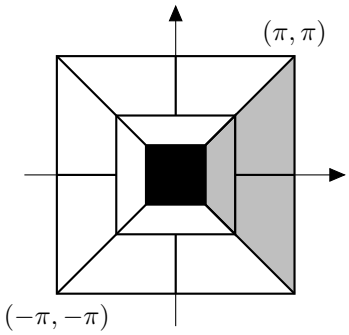


Figure 2.4: Ideal partitioning of the frequency plane of a contourlet transform with 8 directions on the finest high pass level and 4 on the coarsest high pass level, as in Figure 2.2b. The outer square is the frequency plane. Each level of the contourlet transform keeps the center square of the former level. The black center region is contained in the low pass subband and we refer to the gray shaded region as a major direction of the contourlet transform, as these are the directional frequencies covered by the coarsest highpass subband.

introduce a discrete hidden state S_i indicating which normal distribution in the mixture the observation stems from.

- To model the relation between a coefficient o_i and its parent $o_{\rho(i)}$, we model the relation between the associated hidden variables S_i and $S_{\rho(i)}$, and this relation is Markovian, i.e., S_i depends on the rest of the hidden variables only through its parent.

Each coefficient on the coarsest highpass subbands is the root of a tree and each tree is modelled by a hidden Markov model – as illustrated in Figure 2.5.

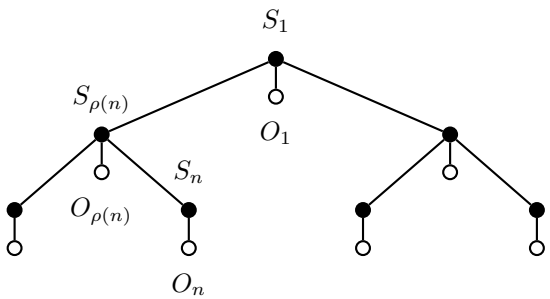


Figure 2.5: Illustration of a graph and the parent-children relations of a hidden Markov model. The black nodes are hidden state variables and the white nodes are observations. Index 1 is the root of the tree and n is an arbitrary node with parent $\rho(n)$.

Let $\mathcal{T}$ be such a tree. If each hidden state take on values $\{1, \ldots, M\}$ (unrelated to the M in (2.2)), we parameterise it as follows:

- The distribution of the hidden state at the root, $P(S_1 = m)$, $m = 1, \ldots, M$.
- The transition probabilities between each hidden state and its parent, $\epsilon_{i,\rho(i)}^{n,m} = P(S_i = n | S_{\rho(i)} = m)$, $m, n = 1, \ldots, M$.
- The mean $\mu_{i,m}$ and standard deviation $\sigma_{i,m}$ of each of the normal distributed variables $O_i | S_i = m$, $m = 1, \ldots, M$.

The free parameters of the tree, which we use later in the classification, removes the redundant probabilities from the list above.

To obtain robust estimates of the parameters in the model, we model the trees rooted in the same major direction (ref. Figure 2.4 and Figure 2.3b) as independent and identically distributed – this is also known as tying [**13**, **34**].

As mentioned in Section 2.3.1, the number of coefficients quadruple at every level, so if for example we apply a three level contourlet transform to a square image with a side length of 1024 pixels there are $1024^2/4^3 = 128^2$ trees.

We fit hidden Markov models to each of the major directions independently.

In our application, the hidden states can each take two values since this provides an adequate description of our data with the minimum number of parameters. Furthermore, we set $\mu_{i,m} = 0$ for all i and m, since this choice also fits data.

Hidden Markov modelling of contourlet transforms was introduced in [**32**], which was motivated by similar modelling of wavelet transforms from [**13**].

Hidden Markov modelling of wavelet transforms have only limited directional separation and each direction is modelled *separately* as illustrated in Figure 2.3a. The modelling of the contourlet transform offers some modelling of inter-directional dependencies, since the major directions of the contourlet transform can spilt into several minor directions as seen in Figure 2.3b.

To verify that the hidden Markov model is indeed applicable for the images we want to model, we compare the empirical distribution of contourlet coefficients in a subband with coefficients simulated from the fitted hidden Markov model. As a by-product of the maximum-likelihood estimation we get the distribution of the hidden states in the hidden Markov model, and therefore simulation from the mixture model in a subband is straightforward. The validity of the model is verified by making a QQ-plot of the observed and simulated coefficients.

The idea of a QQ-plot was introduced in [**40**] as a visual method to compare two distributions and will be explained here.

The distribution of a stochastic variable X is characterized by its cumulative distribution function (cmf), i.e., the function $F_X : \mathbb{R} \to [0, 1]$ defined by $F_X(x) = P(X \leq x)$.

The empirical cumulative distribution function (ecmf) of a sample $x_1, \ldots, x_n$ is the

function $F_n : \mathbb{R} \to [0,1]$ defined by

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{x_i \leq x\}}(x). \tag{2.3}$$

To determine if a sample $x_1, \ldots, x_n$ comes from the same distribution as X, we can compare the cumulative distribution function (cmf) F_X and the ecmf F_n. Unfortunately it is often difficult to decide whether two cmfs are sufficiently equal to conclude that the distributions are alike. The difficulty stems from the fact that it is often more difficult to compare curved graphs than straight lines, especially when the curves are close to 0 or 1.

In a QQ-plot this issue is resolved by linearizing the problem and comparing fractiles of the distributions.

Let $F(x-)$ be the limit from the left of F at x, that is

$$F(x-) = \lim_{z \nearrow x} F(z).$$

Then the fractile function of a cmf F is the function

$$Q : [0,1] \ni p \mapsto \{x \in \mathbb{R} \,|\, F(x-) \leq p \leq F(x)\}. \tag{2.4}$$

A cmf (and thereby the distribution) is characterized by its fractiles and convergence in law implies pointwise convergence of corresponding fractile functions [**24**]. This means that instead of determining if the ecmfs converges to a spcific cmf, we focus on convergence of the corresponding fractiles.

As it is seen from (2.3) an ecmf is piecewise constant and the fractiles i/n, $0 \leq i \leq n$, are not unique. However, the fractiles are determined by the jumps of of F_n, which happens at the lower fractiles, defined by $\breve{Q} : [0,1] \to \mathbb{R}$,

$$\breve{Q}(p) = \inf\{x \in \mathbb{R} \,|\, Q(x) \geq p\}. \tag{2.5}$$

We use Q_n to denote the fractile function of the ecmf F_n and furthermore we let $x_{(1)}, \ldots, x_{(n)}$ be the order variables of $x_1, \ldots, x_n$, which is simply a re-indexing of $x_1, \cdots, x_n$ such that

$$x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}.$$

To compare the empirical distribution to a proposed distribution with fractile function Q, we plot the points

$$(\breve{Q}(i/n), \breve{Q}_n(i/n)) = (\breve{Q}(i/n), x_{(i)}), \quad 1 \leq i \leq n. \tag{2.6}$$

If the samples $x_1, \ldots, x_n$ do indeed follow the proposed distribution, these points will lie on a straight line[1].

In our case, where we want to compare the empirical distribution of the contourlet coefficients, with simulations from the fitted mixture distribution, the function Q in (2.6) is the empirical distribution function from the simulated values.

[1]The coefficients from a contourlet transform are not independent; as indicated in Figure 2.3, coefficients with the same parent are related. When using the dependent observations in a QQ-plot we need to weigh the observations accordingly [**40**].

Parameter estimation

We estimate parameters in the hidden Markov models by maximizing the corresponding likelihood function using an Expectation Maximization (EM) algorithm; the appropriate algorithm for this type of model is described in [**13**]; see also [**34**] for precautions related to the implementation.

The EM algorithm is an iterative procedure and requires an initial estimate of the parameters as input. We initialize by simply choosing a random model obtained by simulating random values for each of the parameters.

To reduce the final model's dependency on the initial estimates, we run the EM algorithm a number of times with random initial values and choose the final model that yields the highest likelihood.

During a maximization step of the EM algorithm, it may happen that some parameters are updated to values that are approaching the computers numerical precision, which is not desirable. We therefore restrict the permissible values of the parameters to avoid numerical underflow. This probably also makes the EM algorithm less sensitive to the initial estimates [**23**].

2.3.3 Classification

Hidden Markov models describe our data adequately and are also widely used for classification purposes, e.g. speech recognition [**34**] and texture retrieval [**13**, **28**, **32**].

Classification is typically performed by computing a sensible distance between the fitted models, e.g. the Kullback-Leibler divergence. For our data this did not give satisfactory results, so we have used the models in a different way.

Using the parameters of the hidden Markov models directly with a support vector machine classifier (SVM; see e.g. [**22**] for a good exposition) was not successful either.

A likely explanation to the failure of the SVMs is that many of the parameters in the hidden Markov models are highly correlated; we circumvent this issue by performing a feature selection prior to the classification.

We have tried two kinds of feature selection that both increased the success of the classifier. SVM can be used to perform recursive feature selection as described in [**21**] by recursively filtering out the features with the smallest weights (as assigned from the SVM).

However, the recursive feature selection was not as successful as using the lasso for logistic regression, see Section 2.3.3. The individual steps will be described in more

details afterwards, but the workflow of our classification is as follows:

$$\begin{aligned}&\text{model parameters}\\ &\xrightarrow{\text{lasso}}\text{ select \& weigh parameters}\\ &\longrightarrow\text{ weighted distances between images}\\ &\xrightarrow{\text{MDS}}\text{ embedding in }\mathbb{R}^2\\ &\xrightarrow{\text{SVM}}\text{ classification}\end{aligned}$$

Parameter selection

In the following we let $\boldsymbol{\theta}_n = (\theta_{n,1}, \ldots, \theta_{n,K})$ denote the free parameters of the hidden Markov model fitted to an image n. Furthermore, we let $\widetilde{\boldsymbol{\theta}}_n = (1, \theta_{n,1}, \ldots, \theta_{n,K})$. We are interested in modelling the paintings as being either authentic or forgeries, i.e., as having two possible states. When using logistic regression to select and rank the parameters, we model this as the probabilities

$$p_n = P(\text{image } n \text{ is authentic}|\boldsymbol{\theta}_n)$$

using the logistic transformation, $x \mapsto \log \frac{x}{1-x}$. The simplest type of regression is when we assume a linear relationship

$$\log \frac{p_n}{1-p_n} = \beta_0 + \sum_{\ell=1}^{K} \beta_\ell \theta_{n,\ell} = \boldsymbol{\beta}^\top \widetilde{\boldsymbol{\theta}}_n. \tag{2.7}$$

There is no a priori justification for using a linear model, except that it is the simplest model and it gives satisfactory results.

Furthermore, the model (2.7) has the advantage that it does not depend on the range of the pixel values of the images we classify. We typically represent the grayscale values of the pixels in a digital image either as floating point numbers in the unit interval or as 8 or 16 bit integers, i.e., integers between 0 and $2^8 - 1$ or $2^{16} - 1$, respectively. The representation range should be consistent in our dataset, but the classification method should not depend on the specific choice. That (2.7) ensures this will be elaborated on after the definition of the metric we introduce to measure the distance between the hidden Markov models.

Under the assumptions of a two class model and (2.7), the log-likelihood for N images under the Bernoulli model is

$$\begin{aligned} l(\boldsymbol{\beta}) &= \sum_{i=1}^{N} \big\{ y_i \log p_i + (1-y_i) \log(1-p_i) \big\} \\ &= \sum_{i=1}^{N} \Big\{ y_i \boldsymbol{\beta}^\top \widetilde{\boldsymbol{\theta}}_i - \log\big(1 + \exp(\boldsymbol{\beta}^\top \widetilde{\boldsymbol{\theta}}_i)\big) \Big\}. \end{aligned} \tag{2.8}$$

It is well known how to maximize this log-likelihood function, see e.g. [**22**].

However, using the regular logistic regression for feature selection gives rise to some problems: 1) it is difficult (and sometimes impossible with the working precision) to estimate the β's in high dimensional problems, and 2) many of the β's have small numerical values, but few are exactly zero, i.e., we do not perform a good selection.

To circumvent this problem, we use the lasso logistic regression [**22**] instead, where we maximize (2.8) subject to the constraint

$$\sum_{\ell=1}^{K} |\beta_\ell| \leq t. \tag{2.9}$$

Equivalently, we can solve the optimization problem in the Lagrangian form

$$\max_{\boldsymbol{\beta}} \Big\{ \sum_{i=1}^{N} \Big[y_i \boldsymbol{\beta}^\top \widetilde{\boldsymbol{\theta}}_i - \log\big(1 + \exp(\boldsymbol{\beta}^\top \widetilde{\boldsymbol{\theta}}_i)\big) \Big] - \lambda \sum_{\ell=1}^{K} |\beta_\ell| \Big\}. \tag{2.10}$$

In [**18**] the authors present a fast iterative algorithm for solving (2.10) and we have used the implementation [**19**]. When using the software from [**19**] we get solutions to (2.10) for finite, decreasing sequence of λ's. The optimal λ is then chosen by leave-one-out cross-validation, as illustrated in Algorithm 2.

Algorithm 2: Cross-validation

Input: patches from N images, where the i'th image consists of N_i patches:
$P = \{p_{i,j} \,|\, 1 \leq i \leq N, 1 \leq j \leq N_i\}$
Input: decreasing, positive sequence $\Lambda = \{\lambda_k\}_{k=1}^{K}$
Output: cross-validation error e_k for each λ_k
Initialization:
$e_k = 1$, $1 \leq k \leq K$
for $i \in \{1, \ldots, N\}$ **do**
  patch set $= P \setminus \{p_{i,j} \,|\, 1 \leq j \leq N_i\}$
  for $\lambda \in \Lambda$ **do**
    train model from (2.10) with patch set and λ
    classify $\{p_{i,j} \,|\, 1 \leq j \leq N_i\}$ with model
    if *classification is correct* **then**
      $e_k := e_k - 1/N$
return $e_1, \ldots, e_K$.

With a set of estimated β's we define the associated weights as

$$w_i = |\beta_i|, \quad 1 \leq i \leq K. \tag{2.11}$$

Using these weights we define the distance between two hidden Markov models with parameters $\boldsymbol{\theta}_i$ and $\boldsymbol{\theta}_j$, respectively, as

$$d(\boldsymbol{\theta}_i, \boldsymbol{\theta}_j) := \sum_{\ell=1}^{K} w_\ell |\theta_{i,\ell} - \theta_{j,\ell}|. \tag{2.12}$$

Since many of the weights are zero, the number of non-zero terms in the sum (2.12) is usually much smaller than K. The reason for using a weighted ℓ^1-norm (as opposed to an ℓ^p-norm with $p > 1$) is that the ℓ^1-norm penalize small differences better than ℓ^p-norms with larger p, which is necessary for the probability parameters. The classification is significantly better when using the ℓ^1-norm instead of the ℓ^2-norm. Using ℓ^τ pseudo-norms with $0 < \tau < 1$ does not further improve the classification.

The metric (2.12) is between the hidden Markov models fitted to the individual patches in an image; we combine the distances between patches into a distance between images by means of the Hausdorff distance: Let N_I denote the number of patches in picture I and let $\boldsymbol{\theta}_{I,1}, \ldots, \boldsymbol{\theta}_{I,N_I}$ be parameters of the hidden Markov models fitted to the patches in I. Then the Hausdorff distance *from* image I to image J is

$$d_H(I, J) := \max_{1 \le k \le N_I} \big\{ \min_{1 \le \ell \le N_J} \{ d(\boldsymbol{\theta}_{I,k}, \boldsymbol{\theta}_{J,\ell}) \} \big\}, \tag{2.13}$$

and the Hausdorff distance *between* image I and J is

$$\max\{d_H(I, J), d_H(J, I)\}. \tag{2.14}$$

Before elaborating on how we use these distances for classification, we shall prove the claim from the beginning of the section, i.e., that these distances do not depend on the range of the pixel values in the images.

The probability parameters in the hidden Markov models do not depend on the range of the pixel values; hence we need only pay attention to the standard deviations. Both the contourlet transform and the standard deviations are positively homogeneous of degree 1 (when scaling the pixel values, the contourlet coefficients scale similarly; the same goes for the deviances of the fitted hidden Markov model) and the β's in the regression are homogeneous of degree -1 (if a parameter $\theta_{n,\ell}$ in (2.7) scales with a factor $k > 0$, the corresponding weight β_ℓ scales with $1/k$). This implies that the scaling factor cancel in the expression (2.12): Let $\boldsymbol{\theta}_i$ and $\boldsymbol{\theta}'_i$ be the parameters of two hidden Markov models fitted to the same image, but with pixel values scaled differently. If $\sigma_{i,\ell}$ and $\sigma'_{i,\ell}$ are standard deviations in $\boldsymbol{\theta}_i$ and $\boldsymbol{\theta}'_i$, respectively, then $\sigma_{i,\ell} = k\sigma'_{i,\ell}$ for some constant $k > 0$ and the corresponding regression weights are $w_\ell = \frac{1}{k} w'_\ell$. Thus the contribution of these standard deviations to the distance (2.12) is

$$w_\ell |\sigma_{i,\ell} - \sigma_{j,\ell}| = \frac{w_\ell}{k} |k\sigma_{i,\ell} - k\sigma_{j,\ell}| = w'_\ell |\sigma'_{i,\ell} - \sigma'_{j,\ell}|,$$

independent of k.

Returning to the use of (2.14), we use (2.12) to (2.14) and compute pairwise distances between all the images. Using an MDS algorithm, we find a configuration of points in $\mathbb{R}^2$ whose pairwise distances are as close as possible to those computed by (2.14).

Here we briefly recall the two approaches to MDS and refer to the literature for details and algorithms; see [**36**] for a thorough exposition on the so-called classic MDS and e.g. [**17**, **22**] about non-classical MDS.

The starting point in MDS is a *dissimilarity*: A dissimilarity is a function $d : \{1, ..., n\} \times \{1, ..., n\} \to \mathbb{R}_+$ where $d(i, i) = 0$ and $d(i, j) = d(j, i)$ for all $i, j \in \{1, ..., n\}$ (this is obviously satisfied for a metric). Henceforth we use the notation d_{ij} instead of $d(i, j)$.

In MDS we seek a configuration of points $\boldsymbol{x}_i \in \mathbb{R}^d$, $i \in \{1, ..., n\}$ such that

$$\|\boldsymbol{x}_i - \boldsymbol{x}_j\| \approx d_{ij} \quad \forall i, j \in \{1, ..., n\}, \tag{2.15}$$

as illustrated in Figure 2.6.

In the non-classical MDS, the configuration is determined by minimizing a goodness-of-fit criterion, $F : \bigtimes_{i=1}^{n} \mathbb{R}^d \to \mathbb{R}_+$, where the dimension d is specified.

There are several often used choices for F [**17**]; we have used the standard criterion from Matlab, which is a normalized square error:

$$F(\boldsymbol{x}_1, ..., \boldsymbol{x}_n) = \frac{\sum_{i<j}(\|\boldsymbol{x}_i - \boldsymbol{x}_j\| - d_{ij})^2}{\sum_{i<j} \|\boldsymbol{x}_i - \boldsymbol{x}_j\|}. \tag{2.16}$$

In the classic MDS we consider the matrix of all dissimilarities $D = [d_{ij}]$. We say that D is euclidean if there exists a configuration of points $\boldsymbol{x}_i$, $i \in \{1, ..., n\}$, in *some* $\mathbb{R}^d$. We have the following result on classic MDS [**36**]: D is euclidean if and only if the matrix

$$B = (I_n - n^{-1}\mathbf{1}_n\mathbf{1}_n^\top)A(I_n - n^{-1}\mathbf{1}_n\mathbf{1}_n^\top)$$

is positive semidefinite ($\mathbf{1}_n$ is the $n \times 1$ vector where all entries are 1 and $^\top$ denotes the transpose). The configuration is also computed from the matrix B and the dimensions of the configuration is ordered after the values of the positive eigenvalues of B. This ordering of the dimensions has practical implications. If the largest eigenvalues are also large compared to sum of the eigenvalues, we can obain a reasonable euclidean configuration in fewer dimensions than where we are promised an exact euclidean configuration. This is also true even if B is not positive semidefinite (meaning that some of the eigenvalues are negative).

In the plots where we show a configuration found by an MDS it is always in $\mathbb{R}^2$ and we label the axes x and y, resectively. As noted below, the units in these plots are not important for the interpretation.

From the points computed by the MDS we construct a classifier based on a simple SVM with linear decision boundary.

The important part of this procedure is that we arrive at a configuration of points where the images of authentic paintings and forgeries are indeed *separable* – the subsequent classification by linear SVM might not be optimal, but it serves the point of illustrating the potential of our method.

It is worth making a couple of remarks.

- In plots where the training is based on different images, the weights from (2.11) are different and hence the pairwise distances between the hidden Markov models in (2.12) change as well. This means that it does not make sense to make direct

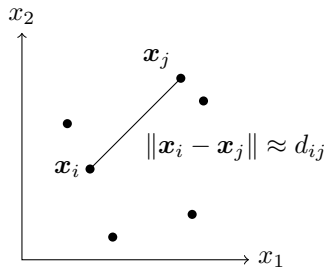


Figure 2.6: Illustration of the purpose of multidimensional scaling: From a set of pairwise dissimilarities d_{ij}, $1 \leq i, j \leq n$, we wish to find a configuration of points $\boldsymbol{x}_i \in \mathbb{R}^d$, $1 \leq i \leq n$, whose pairwise distances resembles the pairwise dissimilarities.

comparison of the subsequent plots in Figure 2.8 to 2.11. Consider Figure 2.10 and 2.11 as an illustration: Here the pairwise distances between the Caspers images change as different weights are used, whereas their spatial distribution does not change significantly.

- For the purpose of visualizing the pairwise distances between the models, the units in the MDS plot are not important – in this regard only the relative distances matters.

  When computing a classification rule with SVM, the points are normalized to have zero mean and unit standard deviation, as there are indications that this ensures a more robust rule [**2**, **20**]. Hence the units are not important for the SVM classification either.

  In conclusion, we do not need units for the classification nor the display of points, but we have included the units to illustrate how the weights influence the distances.

2.4 Results

We now present the results of our experiments with the paintings related to Asger Jorn and Charlotte Caspers.

All numerical experiments were carried out in Matlab and the source code is distributed[2].

We used the Contourlet Toolbox by Minh Do [**14**]. The contourlet transform in [**14**] is implemented with a double filter bank, one that implements a Laplacian pyramid for separating low pass and highpass subbands, and one that implements the partitioning into directional subbands. We therefore need to specify filters for the two filter banks, the number of levels in the Laplacian pyramid, and the number of directional subbands on each level of the Laplacian pyramid.

[2]Available online at `http://www.mathworks.com/matlabcentral/fileexchange/35322`

The classification results with the different filters were comparable and the number of levels and directions had only a small influence. The results presented here are based on the 9-7 pyramidal filter, the pkva12 directional filter and a four level contourlet transform with 8, 8, 16 and 16 directions from coarsest to finest pyramidal level.

A topic, which is not related to the transforms used, is the options considering downsampling of the images. Obviously we cannot use images of arbitrary resolution; if the resolution is too coarse the level of details is not sufficient; if the resolution is too high we may be modelling insignificant details in the canvas or paint because the brushstrokes are not part of high frequency content.

The effect of downsampling – and other preprocessing of the images – will not be investigated thoroughly here; these issues need further investigation and will be dealt with in our subsequent work. In the present experiments we have fitted hidden Markov models and estimated the prediction error of the corresponding classifiers through cross-validation for a variety of downsampling factors and used the downsampling factor that yields the best result.

Our experiments showed that using an inappropriate amount of downsampling is not only visible in the final classification; it may entail that the optimization algorithm used to solve (2.10) is not converging and hence fitting the Bernoulli model is not feasible.

As described in Section 2.3.2, we verify that the HMT model is appropriate for our data by making a QQ-plot of the observed coefficients versus coefficients simulated by the fitted HMT model. An illustration of a QQ-plot for the coefficients is presented in Figure 2.7.

For some subbands the tails of the coefficient distributions are heavier than the fitted mixture distribution - however, the model is acceptable.

2.4.1 Results for Asger Jorn

As mentioned in Section 2.2.1 some of the paintings from Museum Jorn were photographed in multiple images and we divided our data into two groups, one for classification based on the digital images, and one for classification based on the paintings, i.e., by concatenating the images of the same painting. This means that when classifying the digital images we used (2.13) as it stands, but when classifying paintings we calculated the pairwise distances between the paintings P and P' as

$$d_H(P, P') := \max_{1 \le k \le N_P} \Big\{ \min_{1 \le \ell \le N_{P'}} \{ d(\boldsymbol{\theta}_{P,k}, \boldsymbol{\theta}_{P',\ell}) \} \Big\}, \tag{2.17}$$

where N_P is the number of hidden Markov models fitted to the patches in *all* images from P.

The result of the multidimensional scaling and the following classification with a linear SVM is seen in Figure 2.8 and summarized in Table 2.2.

The λ's in the regression model (2.10) that give the best classification results rely on between 10 and 30 parameters from the HMTs – which is a considerable reduction from the full parameter set of 184 parameters.

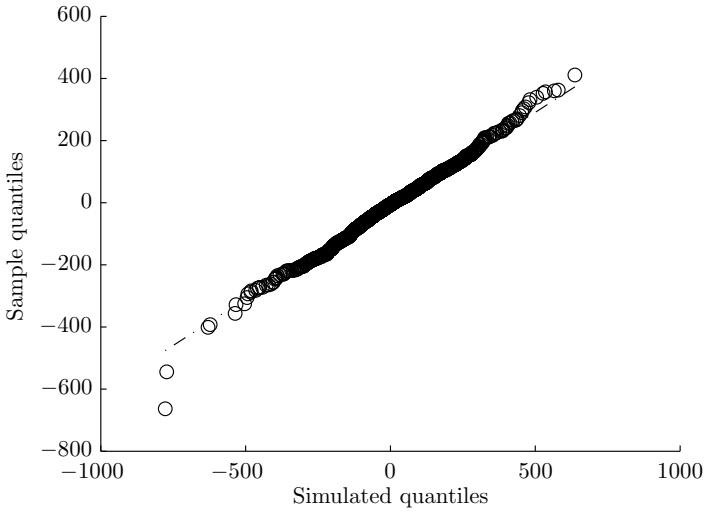
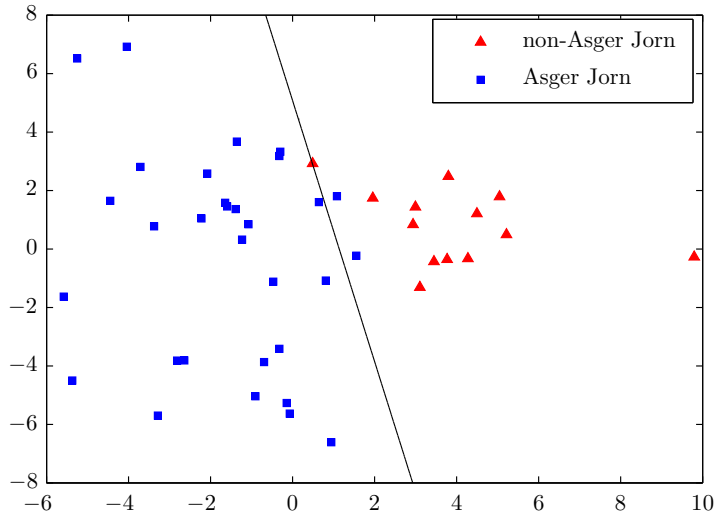


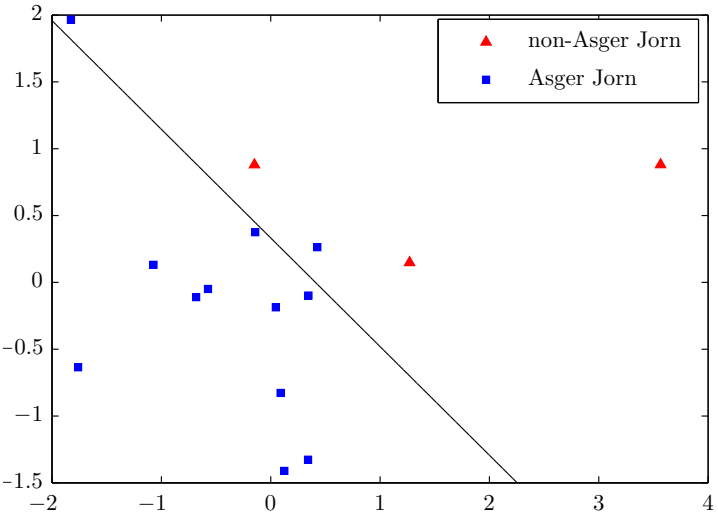
Figure 2.7: QQ-plot of coefficients from subband versus simulated coefficients from the fitted HMT model. The coefficients are from a patch of an Asger Jorn picture recorded with the Nikon D90 camera.

	AJ	not AJ	total
images	28 (31)	11 (13)	39 (44)
paintings	9 (12)	2 (3)	11 (15)

Table 2.2: Classification results of the images related to Asger Jorn. As discussed in the paper some paintings are captured in multiple images. The first number in a category is the number of correctly classified images/paintings; the number in parenthesis is the total number of images/paintings in that category.



(a) Classification of the Asger Jorn images.



(b) Classification of the Asger Jorn paintings. The pairwise distances are smaller than in Figure 2.8a due to the second application of the Hausdorff distance (2.13).

Figure 2.8: Classification of Asger Jorn's images using a linear SVM on the embedding in $\mathbb{R}^2$ computed with a multidimensional scaling algorithm. In Figure 2.8b the classification is based on the paintings – each of which is captured in several images. Figure 2.8a shows the classification of the individual images.

Furthermore, it is notably difficult to classify the images of the painting in Figure 2.1b by Helmut Sturm; when performing cross-validation, the painting by Helmut Sturm is misclassified for all λ values.

As mentioned in Section 2.2.2 other researchers have reported difficulties with classification independent of the camera, i.e., classifying digital reproductions of paintings related to the same artist, but acquired by different cameras/scanners [**33**].

It is of interest to see how our method performs in such a situation and as mentioned in Section 2.2.1 we photographed the paintings at Museum Jorn with two different cameras. Comparing the two sets of images should yield good results since they are digital reproductions of the same paintings. However, since the two cameras produce digital reproductions that differ significantly in resolution and DPI, we need to downsample the high resolution images to a level where the details are of the same size as in the low resolution images. If the high resolution images are not downsampled, the objects in the two sets of images are not comparable in any way. In our situation we can determine the appropriate level of downsampling by comparing the size of a painting's characteristic elements in different images of that painting.

As it is seen in Figure 2.9 our method can perform cross-camera classification if we choose the right level of downsampling; it is also evident that a comparison of images from the two cameras is meaningless if the level of downsampling is inappropriate.

With Figure 2.9b we correctly classified 30 out of 36 images from the Canon camera.

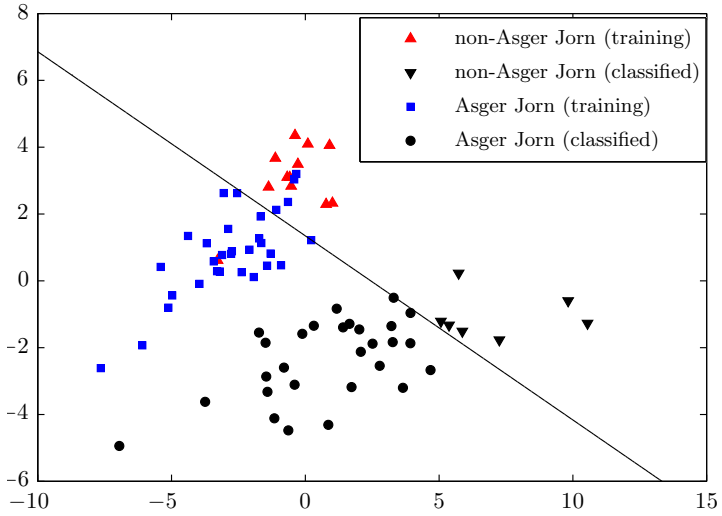
2.4.2 Results for Charlotte Caspers

Our results regarding the images by Charlotte Caspers are far from satisfactory in terms of the number of correct classifications, but they are nonetheless interesting.

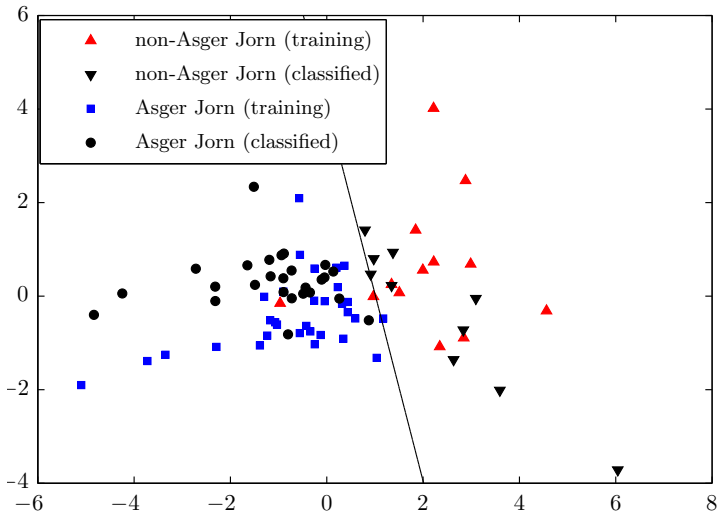
In Figure 2.10 is the result of a MDS for all images by Caspers and here it is worth noticing that the copies and originals are clustered pairwise and separated from the other pairs.

Also, for many λ's in (2.10), the regression algorithm did not converge and in all cases the number of non-zero parameters were very high compared to the experiments with Asger Jorn. This indicates that the Bernoulli model is not suitable, as more than two classes are needed.

It is interesting to study the behaviour of our method when classifying paintings that are totally unrelated to the training material, e.g. the paintings by Charlotte Caspers using a classification rule obtained from the paintings related to Asger Jorn. The result is presented in Figure 2.11; as expected the results related to Charlotte Caspers' paintings do not overlap with those of Asger Jorn.



(a) The zoom level of the training and test images are very different.



(b) The zoom level of the training and test images are comparable.

Figure 2.9: Comparison of the effect of zooming the images. In both cases we have tried to classify the images from the Canon camera (with low resolution) from a Support Vector Machine decision rule calculated from images obtained with the Nikon camera (with different resolutions). In Figure 2.9a we used the original images for the training and in Figure 2.9b we scaled the training images. The right zoom level was determined manually by comparing fixed objects in images of the same painting with different zoom levels.

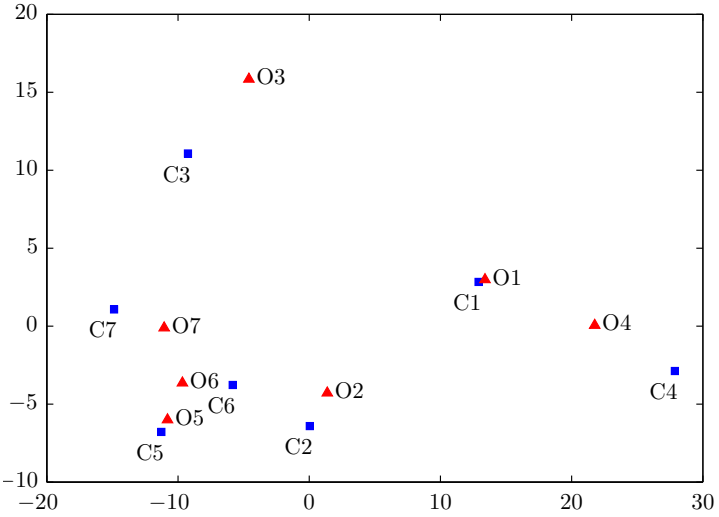


Figure 2.10: Embedding in $\mathbb{R}^2$ of the models fitted to paintings by Charlotte Caspers from a multidimensional scaling algorithm, as described in Section 2.3.3. The originals and copies are denoted by O# and C#, respectively, with # = $1, \ldots, 7$.

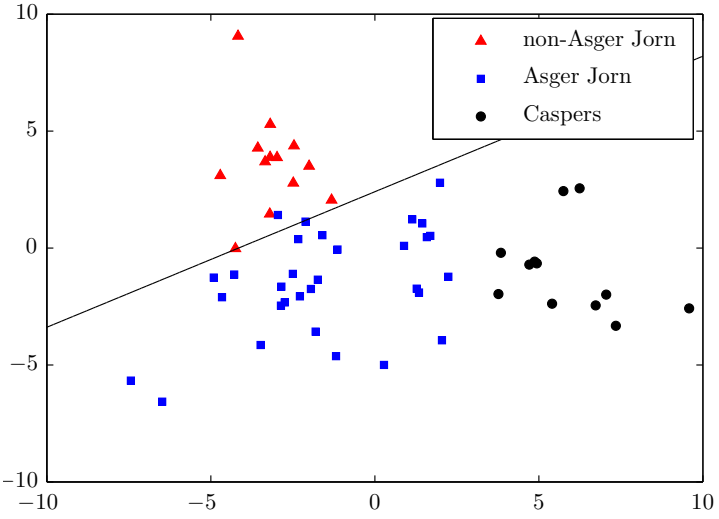


Figure 2.11: Classification of the paintings by Charlotte Caspers using the paintings related to Asger Jorn. As can be seen, the paintings by Charlotte Caspers do not overlap the paintings related to Asger Jorn.

2.5 Discussion

The classification method presented yields good results when separating paintings by Asger Jorn from those of his collaborators; however, there are at least two issues that are worth mentioning: 1) our method *do* contain black-box elements, and 2) the poor classification of the paintings by Charlotte Caspers in Figure 2.10.

Regarding Item 1) it would of course be desirable to know exactly what we are capturing in the paintings using the hidden Markov models – is it indeed the brushstrokes? On the other hand, all present methods for digital authentication contain some kind of black-box element; there is no guarantee that our interpretation of the features is exhaustive.

Regarding Item 2) it is disappointing that the method cannot handle this data set. On the other hand, the "copies" of Charlotte Caspers' paintings are made by herself and thus the styles can be expected to have a high degree of similarity – as can be seen online [**9**]. Hence it may be that the reason that our method indicates the pairwise cohesion is because it is in fact so. This conjecture is supported by the observation from Section 2.4.2 of the Bernoulli model being inappropriate for modelling the hidden Markov models: The originals and copies are not two homogeneous classes.

This may be expanded to a wider generality. When training a classifier to distinguish between the works of a given artist and "the rest", it is not realistic to model all of "the rest", since "the rest" potentially contains every image not made by the given artist, which is definitely not homogeneous.

In the situation where we have multiple imitations that could not be modelled as representations of the same class, one could imagine training multiple classification rules – one for each of the homogeneous subclasses – and deem a painting authentic if it passed all/a majority of these classification rules.

Finally, it is interesting that our classification methods depend on the zoom level of the images. The effect of data collection and preprocessing of the digital images is an issue that will be addressed further in later work.

To compare our method with the traditional hands on approach we asked curator at Museum Jorn, Teresa Østergaard Pedersen, to quantify what features in Jorn's paintings she finds distinctive. Besides the descriptions she also mentioned that determining if a painting is made by Asger Jorn takes about one week. As mentioned in Section 2.1 the conclusions may be subjective, but it gives us a sound starting point. The following is based on personal communication with Teresa Østergaard Pedersen as well as the cited references.

A introduction to Asger Jorn's career can be found in [**6**, **10**], but an important perspective in the authentication business is the development of his painting style. In 1948 Asger Jorn co-founded the COBRA group [**11**] which was group of expressionist painters that introduced a more spontaneous abstract art. Jorn also came to believe that paintings should not be too photographic, so unlike his COBRA fellows, Asger Jorn has figure-*like* elements with many shapes and colors; the paintings have a visual balance and entire canvas is used, concretely as well as

metaphorically; a painting does not consist of many different parts, but instead of one visual unit.

The COBRA group dissolved in 1951, but afterwards Asger Jorn's work tended to have more of a red line than before. Importantly for our focus is that his earlier works had larger homogeneous parts and that this is often what characterizes forgeries. Furthermore, the overall impression of the painting is important; often more so than the details.

Among the characteristics of Asger Jorn's style is that he used very thick paint (also referred to as impasto [**12**]) and combined with a finalizing plop of the brush the splashed paint makes the paintings slightly three dimensional in many areas. Furthermore, there are almost no vertical or horizontal lines in the paintings.

In conclusion, there are both pros and cons for using our method in connection with Asger Jorn. The most significant pro is that Asger Jorn avoided the vertical and horizontal lines, thereby confirming the need for a variety of directions in our multiresolution transform.

A con that goes against all methods developed for digital artist authentication is that traditional assessments of paintings related to Asger Jorn are based on holistic impressions.

2.6 Conclusion

In this paper we have constructed methods for classifying paintings. The contourlet transform of digital images of paintings is able to efficiently capture the contours of the image – including the brushstrokes of a painter, which is believed to be characteristic. The contourlet transform is well described by a hidden Markov model, which captures both the coefficient distribution on the individual resolutions and the dependency structure between the resolutions.

By applying a lasso regression we perform variable selection and weighing among the features of the hidden Markov models. Using such a selection to define a metric between the trained hidden Markov models, we find an embedding of points in $\mathbb{R}^2$ that resembles the pairwise distances. Once this embedding is found, we train a classifier to distinguish between the classes of paintings.

With this procedure we have performed a leave-one-out cross-validation where we have successfully classified 39 out of 44 images.

A promising aspect of our method is that we can determine the authenticity of images digitized differently than the training data – from the model found through the cross-validation mentioned above, 30 out of 36 images from the other test set was classified correctly.

Acknowledgement

The authors would like to thank Arne Jensen (Department of Mathematical Sciences, Aalborg University) for suggesting this topic and for providing contact to the Portinari Project, that initiated our interest in this subject. Furthermore, we thank Museum Jorn, Silkeborg, Denmark, and curator Teresa Østergaard Pedersen for allowing us to use paintings of Asger Jorn.

Finally, we would like to thank P. Svante Eriksen (Department of Mathematical Sciences, Aalborg University) for helpful discussions and valuable comments on this manuscript as well as the anonymous reviewers for their valueable feedback.

References

[1] Mahmoud Al-Ayyoub, Mohammad T. Irfan, and David G. Stork. "Boosting multi-feature visual texture classifiers for the authentication of Jackson Pollock's drip paintings". In: *Proceedings of SPIE. Computer Vision and Image Analysis of Art II*. Ed. by David G. Stork, Jim Coddington, and Anna Bentkowska-Kafel. Vol. 7869. 2011.

[2] Shawkat Ali and Kate Smith-Miles. "Improved Support Vector Machine Generalization Using Normalized Input Space". In: *AI 2006: Advances in Artificial Intelligence*. Ed. by Abdul Sattar and Byeong ho Kang. Vol. 4304. Lecture Notes in Computer Science. Springer Berlin/Heidelberg, 2006, pp. 362–371. ISBN: 978-3-540-49787-5. DOI: 10.1007/11941439_40.

[3] Igor Berezhnoy, Eric Postma, and H. Jaap van den Herik. "Authentic: Computerized Brushstroke Analysis". In: *IEEE International Conference on Multimedia and Expo*. 2005. DOI: 10.1109/ICME.2005.1521739.

[4] Igor Berezhnoy, Eric Postma, and H. Jaap van den Herik. "Computer analysis of Van Gogh's complementary colours". In: *Pattern Recognition Letters* 28.6 (2007), pp. 703–709. DOI: 10.1016/j.patrec.2006.08.002.

[5] Igor E. Berezhnoy, Eric O. Postma, and H. Jaap van den Herik. "Automatic extraction of brushstroke orientation from paintings". In: *Machine Vision and Applications* 20.1 (Jan. 2009), pp. 1–9. DOI: 10.1007/s00138-007-0098-7.

[6] Hugh Brigstocke, ed. *The Oxford Companion to Western Art*. Oxford University Press, 2001.

[7] Emmanuel J. Candès and David L. Donoho. "Continuous curvelet transform - II. Discretization and frames". In: *Applied and Computational Mathematics* 19 (Sept. 2005), pp. 198–222. DOI: 10.1016/j.acha.2005.02.004.

[8] Emmanuel J. Candès and David L. Donoho. "New Tight Frames of Curvelets and Optimal Representations of Objects with Piecewise C^2 Singularities". In: *Communications on Pure and Applied Mathematics* 57 (Feb. 2004), pp. 219–266. DOI: 10.1002/cpa.10116.

[9] Charlotte Caspers. *Caspers Data Set*. URL: http://www.math.princeton.edu/ipai/datasets.html.

[10] Ian Chilvers, ed. *The Oxford Dictionary of Art*. Oxford University Press, 2004.

[11] Ian Chilvers and John Glaves-Smith, eds. *A Dictionary of Modern and Contemporary Art*. Oxford University Press Inc., 2009.

[12] Michael Clarke and Deborah Clarke, eds. *The Concise Oxford Dictionary of Art Terms*. Oxford University Press Inc.

[13] Matthew S. Crouse, Robert D. Nowak, and Richard G. Baraniuk. "Wavelet-Based Statistical Signal Processing Using Hidden Markov Models". In: *IEEE Transactions on Signal Processing* 46.4 (Apr. 1998), pp. 886–902. DOI: 10.1109/78.668544.

[14] Minh N. Do. *Contourlet Toolbox*. URL: http://www.mathworks.com/matlabcentral/fileexchange/8837.

[15] Minh N. Do and Martin Vetterli. "The Contourlet Transform: An Efficient Directional Multiresolution Image Representation". In: *IEEE Transactions on Image Processing* 14.12 (Dec. 2005), pp. 2091–2106. DOI: 10.1109/TIP.2005.859376.

[16] David L. Donoho. "Sparse Components of Images and Optimal Atomic Decompositions". In: *Constructive Approximation* 17.3 (2001), pp. 353–382. DOI: 10.1007/s003650010032.

[17] Richard O. Duda, Peter E. Hart, and David G. Stork. *Pattern Classification*. 2nd ed. Wiley, 2001.

[18] Jerome Friedman, Trevor Hastie, and Rob Tibshirani. "Regularization Paths for Generalized Linear Models via Coordinate Descent". In: *Journal of Statistical Software* 33.1 (Feb. 2010), pp. 1–22. URL: http://www.jstatsoft.org/v33/i01.

[19] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. *GLMnet for Matlab*. URL: http://www-stat.stanford.edu/~tibs/glmnet-matlab.

[20] Arnulf B. A. Graf, Alexander J. Smola, and Silvio Borer. "Classification in a normalized feature space using support vector machines". In: *IEEE Transactions on Neural Networks* 14.3 (May 2003), pp. 597–605. ISSN: 1045-9227. DOI: 10.1109/TNN.2003.811708.

[21] Isabelle Guyon, Jason Weston, Stephen Barnhill, and Vladimir Vapnik. "Gene Selection for Cancer Classification using Support Vector Machines". In: *Machine Learning* 46 (1 2002), pp. 389–422. DOI: 10.1023/A:1012487302797.

[22] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. 2nd ed. Springer, 2008. ISBN: 978-0-387-84857-0.

[23] Richard J. Hathaway. "Constrained maximum-likelihood estimation for a mixture of M univariate normal distributions". PhD thesis. Rice University, 1983.

[24] Jørgen Hoffmann-Jørgensen. *Probability with a view toward statistics*. Vol. I. Chapman & Hall/CRC, 1994.

[25] James M. Hughes, Daniel J. Graham, and Daniel N. Rockmore. "Quantification of artistic style through sparse coding analysis in the drawings of Pieter Bruegel the Elder". In: *Proceedings of the National Academy of Sciences* 107 (2009), pp. 1279–1283. DOI: 10.1073/pnas.0910530107.

[26] James M. Hughes et al. "Empirical Mode Decomposition Analysis for Visual Stylometry". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2012). DOI: 10.1109/TPAMI.2012.16.

[27] Mohammad Irfan and David G. Stork. "Multiple visual features for the computer authentication of Jackson Pollock's drip paintings: Beyond box-counting and fractals". In: *Proceedings of SPIE*. Image Processing: Machine Vision Applications II. Ed. by Kurt S. Niel and David Fofi. Vol. 7251. 2009. DOI: 10.1117/12.806245.

[28] C. Richard Johnson, Jr et al. "Image Processing for Artist Identification". In: *IEEE Signal Processing Magazine* 25.4 (2008), pp. 37–48. DOI: 0.1109/MSP.2008.923513.

[29] Jia Li and James Z. Wang. "Studying Digital Imagery of Ancient Paintings by Mixtures of Stochastic Models". In: *IEEE Transactions on Image Processing* 13 (2004), pp. 338–351. DOI: 10.1109/TIP.2003.821349.

[30] Jia Li, Lei Yao, Ella Hendriks, and James Z. Wang. "Rhythmic Brushstrokes Distinguish van Gogh from His Contemporaries: Findings via Automated Brushstroke Extraction". In: *IEEE Transactions on Pattern Analysis & Applications* (Oct. 2011). DOI: 10.1109/TPAMI.2011.203.

[31] Siwei Lyu, Daniel Rockmore, and Hany Farid. "A digital technique for art authentication". In: *Proceedings of the National Academy of Sciences* 101.49 (Dec. 2004), pp. 17006–17010. DOI: 10.1073/pnas.0406398101.

[32] Duncan D.-Y. Po and Minh N. Do. "Directional Multiscale Modeling of Images using the Contourlet Transform". In: *IEEE Transactions on Image Processing* 15.6 (June 2006), pp. 1610–1620. DOI: 10.1109/TIP.2006.873450.

[33] Güngör Polatkan et al. "Detection of Forgery in Paintings Using Supervised Learning". In: *IEEE International Conference on Image Processing*. 2009. DOI: 10.1109/ICIP.2009.5413338.

[34] Lawrence R. Rabiner. "A tutorial on hidden Markov models and selected applications in speech recognition". In: *Proceedings of the IEEE* 77 (Feb. 1989), pp. 257–286. DOI: 10.1109/5.18626.

[35] Robert Sablatnig, Paul Kammerer, and Ernestine Zolda. "Hierarchical classification of paintings using face- and brush stroke models". In: *Proceedings of the Fourteenth International Conference on Pattern Recognition*. Vol. 1. Aug. 1998, pp. 172–174. DOI: 10.1109/ICPR.1998.711107.

[36] G. A. F. Seber. *Multivariate Observations*. Wiley, New York, 1984.

[37] David G. Stork. "Learning-based authentication of Jackson Pollock's paintings". In: *Electronic Imaging & Signal Processing* (May 27, 2009). DOI: 10.1117/2.1200905.1643.

[38] Richard P. Taylor, Adam P. Micolich, and David Jonas. "Fractal analysis of Pollock's drip paintings". In: *Nature* 399 (June 1999), p. 422. DOI: 10.1038/20833.

[39] R.P. Taylor et al. "Authenticating Pollock paintings using fractal geometry". In: *Pattern Recognition Letters* 28.6 (Apr. 2007). Pattern Recognition in Cultural Heritage and Medical Applications, pp. 695–702. DOI: 10.1016/j.patrec.2006.08.012.

[40] M. B. Wilk and R. Gnanadesikan. "Probability Plotting Methods for the Analysis of Data". In: *Biometrika* 55.1 (Mar. 1968), pp. 1–17. JSTOR: 2334448.

CHAPTER 3

Robustness of digital artist authentication

Publication details

Co-authors: Morten Nielsen*

* *Department of Mathematical Sciences*
Aalborg University

Journal: Submitted to EURASIP Journal on Image and Video Processing

This is an extended version of the submitted paper with more artistic considerations and elaboration of technical details.

Abstract:
In many cases it is possible to determine the authenticity of a painting from digital reproductions of the paintings; this has been demonstrated for a variety of artists and with different approaches. All the approaches regarding digital artist authentication have focussed on *potential*, while the *robustness* has not been considered, i.e., the degree to which the data collection process influences the conclusion of the method. However, in order for an authentication method to be successful in practice, it needs to be robust to plausible error sources from the data collection.

In this paper we investigate the robustness of the newly proposed authenticity method introduced by the authors based on second generation multiresolution analysis. This is done by modelling a number of realistic factors that may occur in the data collection.

Keywords
Artist authentication, data acquisition, error sources.

3.1 Introduction

In recent years there has been an increasing interest in detecting fake paintings by analyzing digital reproductions of the paintings. Different approaches from different authors include Al-Ayyoub, Irfan, and Stork [**2**], Berezhnoy, Postma, and Herik [**3**], Hughes, Graham, and Rockmore [**14**], Irfan and Stork [**15**], Johnson, Jr et al. [**17**], Li, Yao, Hendriks, and Wang [**18**], Lyu, Rockmore, and Farid [**19**], Polatkan et al. [**22**] and in [**16**] we present yet another method. The approaches have been tested on a variety of painters, among these Vincent van Gogh, Pieter Bruegel the Elder, Jackson Pollock and Asger Jorn. Many of the approaches rely on the assumption that the brushstrokes in a painting or the lines in a drawing are characteristic for its artist – of which Vincent van Gogh due to his very visible brushstrokes is a very good example [**1**, **3**, **4**, **18**, **20**]. As to drawings the way of finishing lines is often said to be characteristic for an artist (personal communication with Niels Borring, National Gallery of Denmark).

Despite the individual successes in the abovementioned papers, we cannot claim to have a general method for detecting forgeries. Especially the article [**22**] gives rise to contemplation when concluding that a novel approach is successful; in [**22**] it is concluded that the camera used to photograph the paintings can heavily influence the classification due to differences in the digitization.

When testing a forgery detection method one should not ignore the origin of the data. This is also important because we do not (yet) have a standard for digital representation of visual art, since we do not yet know the optimal way of acquiring digital images that captures the parts relevant for detecting stylistic differences.

When presenting our method in [**16**] we demonstrated that it was robust against a change between two cameras, meaning that we can classify images photographed by one camera based on images photographed with another camera.

In the present work we look further into the robustness of our method. Along the

way we discuss a number of reasonable factors in the data acquisition that could influence the classification methods.

The motivation for our work has been the following questions:

1. How robust is our method to data collection?

2. Which error sources (if any) influence the decisions of our method significantly?

3. Can we digitally correct the errors induced by the data collection?

A classification method is of limited interest if it necessitates that the digital images are obtained with exactly the same equipment under identical circumstances. Different museums have different photographic equipment and the equipment is replaced over time; indeed, the reason for the discovery of the digitization's influence in [**22**] was due to the use of a new scanner at the van Gogh museum. Hence, we should allow for some fluctuations in photographic equipment, quality and operator skill.

To our knowledge, the effects of image quality and acquisition have not been studied in the context of digital artist classification. We believe, however, that including considerations of this kind is a necessary part of testing the usefulness of the classification method.

In this work we have studied two significant groups of error sources (factors) that one is likely to encounter when photographing paintings, namely position of the camera in relation to the painting, and the properties of the camera.

Let us elaborate on the factors we have chosen to study.

We think it reasonable to assume that a photographer would choose to use a tripod when photographing a painting, thus ensuring that the photograph is not blurred due to camera motion. However, it may be difficult to ensure that the camera and painting are perfectly aligned – touching a valuable painting is usually not an option and often the painting is tilted slightly from the wall. This has led us to consider the influence of the first two factors.

(1) Rotation of the camera compared to the painting.

(2) Keystone effect.

The camera may be rotated compared to the paintings and our method is not a priori rotation invariant.

The keystone effect is present if the camera is not parallel to the painting. If we have keystone effects straight lines in the painting remain straight, but parallel lines do not remain parallel.

Different cameras digitize images differently; as mentioned above, we demonstrated in [**16**] that our method is robust to this. There are, however, a number of other factors related to the recording that we find interesting to investigate. Specifically we have decided to look at the following factors.

③ Distortion caused by the lens.

④ Exposure.

⑤ Sharpening the picture digitally.

⑥ Contrast in the image.

Virtually all lenses cause distortion when the magnification by the lens depends on the distance to the optical axis [**8**]. Distortion does not decrease image quality like other aberrations, but instead it affects the shape of an image [**24**]. The distortion increases when moving further from the center of the image, so when we divide an image in to multiple patches, the lines curve differently in different patches, as illustrated in Figure 3.3.

When recording a photograph, the exposure heavily influences the visual appeal of the final photograph; if the exposure is too high the image is too bright in a non-homogeneous way and conversely the image is darker if the exposure is too low.

Both sharpening and increasing contrast makes the edges in an image stand out more clearly. Since brushstrokes are very fine edges in a painting, the classification might benefit from enhancing those.

The rest of the paper is organized as follows: In Section 3.2 we present the data used in our experiments; in Section 3.3 we briefly recall our classification method from [**16**] and introduce the methods we use in this paper to test the robustness of our method. In Section 3.4 we present the results of our experiments, and finally we draw conclusions in Section 3.5.

3.2 Data

In [**14**, **19**] the authors used images of drawings by Bruegel the Elder to test their classification methods. By courtesy of Daniel Rockmore of Dartmouth College, we have been able to test our methods on the same set of Bruegel drawings. With these images we get a clear separation between the authentic Bruegel images and the known imitations, and therefore we have chosen this data set for the current analysis.

The original Bruegel drawings stem from the Metropolitan Museum of Art in New York and in the following we refer to the drawings by their Metropolitan Museum of Art catalogue number, as it is done in [**19**]. The image catalogue numbers and their categories are summarized in Table 3.1.

The images consist of eight authenticated drawings by Bruegel and five acknowledged Bruegel imitations. In Figure 3.1 we have included an example of an authentic Bruegel drawing.

All images are 16 bit RGB color images recorded in lossless PNG format [**14**, **19**]. In [**16**] we demonstrated that our method can classify images in JPEG format from

Catalog no.	Title	Artist
3	Pastoral Landscape	Bruegel
4	Mountain Landscape with Ridge and Valley	Bruegel
5	Path through a Village	Bruegel
6	Mule Caravan on Hillside	Bruegel
9	Mountain Landscape with Ridge and Travelers	Bruegel
11	Landscape with Saint Jermove	Bruegel
13	Italian Landscape	Bruegel
20	Rest on the Flight into Egypt	Bruegel
7	Mule Caravan on Hillside	—
120	Mountain Landscape with a River, Village, and Castle	—
121	Alpine Landscape	—
125	Solicitudo Rustica	—
127	Rocky Landscape with Castle and a River	Savery

Table 3.1: References for the images used in our analysis and their category as authentic Bruegel drawings or forgeries. A "—" means that the artist is unknown.

Figure 3.1: Example of an authentic drawing by Pieter Bruegel the Elder used in our experiments, namely the image 3 by the Metropolitan Museum of Art catalogue number. The framed patch is the part we use for demonstrating the effects of the factors in later figures.

a classifier trained on images in TIFF format, but this is not a concern here when all images are of the same quality without compression artifacts.

For our analysis all of the images were converted to grayscale using Matlab's rgb2gray command, which combines the three color channels of an RGB image into a grayscale version G by

$$G = 0.299R + 0.587G + 0.114B.$$

The majority of our work, including the image enhancements, has been carried out in Matlab and the code is distributed[1]. However, changing the exposure of the images has been done using Apple Aperture, version 3.1.2. We will elaborate on this in Section 3.3.2.

Note that all our corrections have been performed digitally, as we have not had access to the paintings ourselves. Testing the robustness of our method on images with simulated errors is a proxy of real world errors; on the other hand it allows more controlled experiments, which is preferable to get (preliminary) insight into the effects of errors.

3.3 Methods

3.3.1 Classification method

Here we briefly introduce our classification method from [**16**], which has been used for testing the impact of the different factors; the reader is referred to this paper for a more thorough description.

The basis for our classification method is to work with the contourlet transform [**6**] of the digital images. In short, the contourlet transform is designed to provide a multiresolution representation of digital images with a *user specified* number of highpass orientations. The choice of the contourlet transform as the starting point for our classification gives greater directional high frequency information and there are no issues with discretising a continuous transform.

The highpass subbands of the contourlet transforms contain among others the brushstrokes in a painting at a variety of orientations and resolutions. It is these subbands we model to obtain a classifier between authentic paintings and forgeries.

Statistical modelling of contourlet transforms for classification/recognition purposes is obtained using hidden Markov trees; they have been used extensively for recognition purposes and was introduced for the contourlet transform in [**21**].

The idea is to model the highpass subbands in the contourlet transform that are related through the multiresolution analysis; this means that subbands at different resolutions are related if they have the same preferred orientations. When fitting a hidden Markov tree to a branch of subbands we investigate how the *structure* of the contourlet coefficients evolves as the resolution increases. This is an important

[1] Available online at `http://www.mathworks.com/matlabcentral/fileexchange/35322`

distinction from object recognition in images – we are not interested in matching specific objects in the images, but rather the distribution of details.

The hidden Markov trees used are parametric models. There are numerous parameters in the hidden Markov trees used, and closed form solutions for the maximum likelihood estimates do not exist. Instead the parameter estimation is performed using an Expectation Maximization (EM) algorithm. The EM algorithm is an iterative procedure for performing maximum likelihood estimation, and as such it needs an initial parameter estimate; this also means that the final parameter estimate depends on the initial estimate. To make our method less dependent on initial parameter estimates, we randomly generate a number of initial parameters, run the EM algorithm for each of them, and choose the final model with the highest likelihood.

As a proxy of distance between paintings, we calculate a distance between the hidden Markov trees fitted to the paintings.

The distances between hidden Markov models are calculated as a weighted ℓ^1-norm of their parameters, where the weights are determined according to their impact on the classification of a suitable set of training images.

The basis of the classification is this proxy of pairwise distances between paintings.

After finishing the work in [**16**] we came up with a simpler way of utilizing the pairwise distances for classification.

The classifications in the present work have been performed using a k nearest neighbour method. That is, we classify a test image by a majority vote amongst the k training images with the smallest distance to the test image. A small difference from the usual k nearest neighbour algorithm is that we work *directly* with the distances between the hidden Markov models and not on an embedding of the distance via e.g. multidimensional scaling.

There are some things to keep in mind in this scenario.

- The small number of training images available causes restrictions to the number of neighbours available for classification, and when using only few neighbours, the variance is higher [**12**].
- The nearest neighbour of a point to be classified can be very far away – further than the distance between any training images. This is especially the case if the image to be classified is a painting that looks completely different from the training material.

Besides the classification with a nearest neighbour procedure we can visualize the images and their interrelation by finding an embedding of points in $\mathbb{R}^2$ that respects their pairwise distances – this is obtained using a multidimensional scaling algorithm as described in Section 2.3.3.

In our experiments, where we have to perform a large number of classifications, we need an automatic procedure. However, when confronted with the task of classifying a new image it would be unwise to rely solely on the automatic classification – among others due to the reasons listed above.

3.3.2 Current experiments

To estimate the impact of the factors considered we have used the original data set described in Section 3.2 as the basis for simulating images with the desired flaws.

We will now describe the procedures used to simulate the effect of the individual factors and the specific settings.

(1) Rotation of the camera is simulated by simply rotating the image. This introduces blank corners and a slightly smaller image after cropping. We have rotated the image at integer degrees between -4 and 4.

(2) Keystone effect is a special kind of projective transform, identified by four control points [**13**]. The idea is illustrated in Figure 3.2 – each of the corners in the large square is mapped to the closest point in the smaller quadrilateral, which is one of the control points. In our experiments the control points are determined by how much they deviate from the side. For the keystone correction we have simulated ten uniformly distributed control points. The points were then normalized such that the control point deviating the most was 2.5% of the side length.

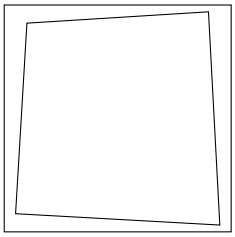


Figure 3.2: Example of keystone effect. Each of the corners in the large square is being mapped to the closest point in the smaller quadrilateral.

(3) To introduce the model used for simulating distortion, we let r_u and r_d denote the centralized polar coordinates in the undistorted and distorted image, respectively.

A standard model for distortion in a lens is Brown's model [**5**], where the relation between r_u and r_d is described by a Taylor series. In many cases only the low order terms for radial distortion are significant [**23**] and we have used the simplest case where

$$r_d = r_u + a \cdot r_u^2.$$

The sign of a determines the nature of the distortion; a negative a gives pincushion distortion and a positive a gives barrel distortion. For the distortion we used the values $0, \pm 1 \cdot 10^{-5}, \pm 1 \cdot 10^{-4}$ for a.

The effect of distortion on a checkerboard pattern is seen in Figure 3.3.

(4) Exposure is difficult to simulate and here we have employed a black box in the form of Apple Aperture. In Aperture we adjust the exposure relative to the

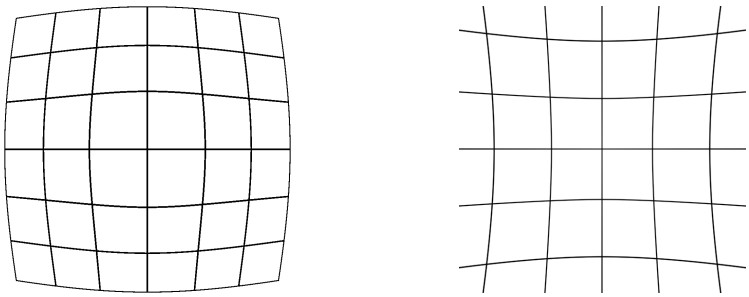


Figure 3.3: Illustration of barrel distortion (left) and pincushion distortion (right).

current value, where a lower exposure darkens the image and a higher exposure brightens the image. We have adjusted the exposure with $\pm 0.5, \pm 0.25, 0.75$ and left it untouched at 0. The visual appearance of the different exposure values is illustrated in Figure 3.4.

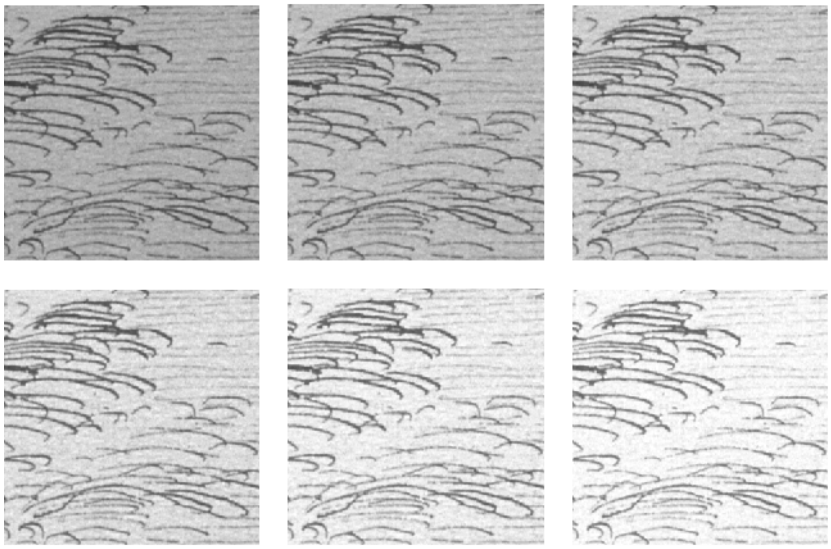


Figure 3.4: Examples from the image 3 with the exposure adjusted. From the left to right the exposure correction is -0.5, -0.25, 0 (original image) in the top row and 0.25, 0.5, 0.75 in the bottom row.

(5) A standard way of sharpening an image is to subtract a blurred version of the same image [**9**], which enhances the edges. Let I, I_s and $\overline{I}$ denote the original, sharpened and blurred image, respectively. The relation is

$$I_s(x, y) = I(x, y) - \overline{I}(x, y)$$

for every pixel (x, y) in I. The blurry image $\overline{I}$ is often computed by applying a discrete differentiation operator on I. In Matlab a 3×3 discrete Laplacian operator is used for blurring, i.e., $\overline{I}(x, y) = \Delta_\alpha I(x, y)$, where α determines the shape of Δ_α by weighing the directions (see [**10**] for details).

We have used no sharpening and the Matlab default with $\Delta_{0.2}$.

(6) Changing the contrast amounts to specifying a transformation of the intensity of the image. The intensity varies between low and high (corresponding to dark and light) and this can be represented as the unit interval. Changing the contrast corresponds to specifying a function $c : [0, 1] \to [0, 1]$ which is surjective.

Initially we increased the contrast of an image by saturating a percentage of the most extreme pixels, which is a standard procedure [**10**]. When saturating $2p$ percentage, the contrast function is given by

$$c(x) = \begin{cases} 0, & x \leq p, \\ \dfrac{x - p}{1 - 2p}, & p \leq x \leq 1 - p, \\ 1, & x \geq 1 - p. \end{cases} \tag{3.1}$$

This way of increasing the contrast is a non-invertible transform and it is illustrated in Figure 3.5. For the saturation we discarded 0, 2, 4, and 6 % of the most extreme pixels.

As explained in Section 3.4 we later decided to apply invertible transforms for the contrast values. Our choice in this regard was the family of functions

$$c_\gamma(x) = x^\gamma, \tag{3.2}$$

where $\gamma > 0$. When $\gamma < 1$ the low intensity parts of the image are easier to distinguish, and when $\gamma > 1$ the high intensity parts are easier to distinguish, as illustrated in Figure 3.5.

As mentioned above we think it natural to divide the factors into two groups – namely those related to the position of the camera, (1) and (2), and those related to the recording properties of the camera, (3) to (6).

To test the influence of the factors we perform a factorial experiment for each of the two groups. For the position experiment we have a total of $9 \cdot 10 = 90$ combinations and for the camera experiment there is a total of $5 \cdot 6 \cdot 2 \cdot 4 = 240$ combinations.

The final goal of a classification method is to have a high rate of correct classifications, hence when making conclusions about the impact of a particular factor, the most important question is whether or not the correct classification rate is significantly downgraded. We estimate this in two ways, namely by performing leave-one-out cross-validation and by classifying the simulated error images with the original images.

The reason for including the second classification method is that all the effects we apply to our images are *relative* to the original data set, so if a factor has a

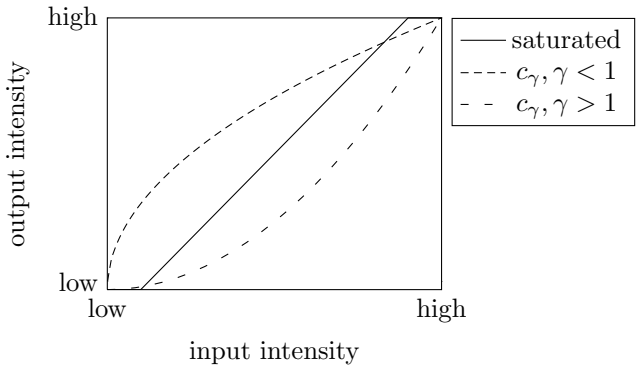


Figure 3.5: Change of contrast in an image is performed by specifying a transformation of the intensity values. Here we show the shapes of typical contrast curves as defined in (3.1) and (3.2).

negligible effect we should be able to classify the generated images correctly with the original images.

As mentioned in Section 3.3.1 our model is fitted using an iterative procedure that relies on an initial parameter estimate, which results in a final model that also depends on the initial estimate. Therefore, we cannot expect different runs to yield exactly the same final model. In order to make conclusions about the effects of the factors studied, we therefore need to get insight into the effects of the EM algorithm.

This is done by training a number of models on the original images and see how they behave compared to one reference model.

This experiment gives rise to some interesting questions:

1. Is the variation caused by the EM algorithm homogeneous no matter which painting?

2. Is the variation caused by the EM algorithm so great that it effects the classification?

3. Is the variation by the factors (1) to (6) greater than what can be expected from the EM algorithm?

3.4 Results

The images we used as training data for the classification scheme are well separated: When performing leave-one-out cross-validation 12 of the 13 images are classified correctly using both 1 and 3 nearest neighbour classification; neither case can classify the image 125 correctly.

From the pairwise distances used for classification we have also computed embeddings as mentioned in Section 3.3; the result of this embedding is shown in Figure 3.6.

From Figure 3.6 it is seen that image 125 is misclassified because it is far from *both* the authentic training images and the other known forgeries.

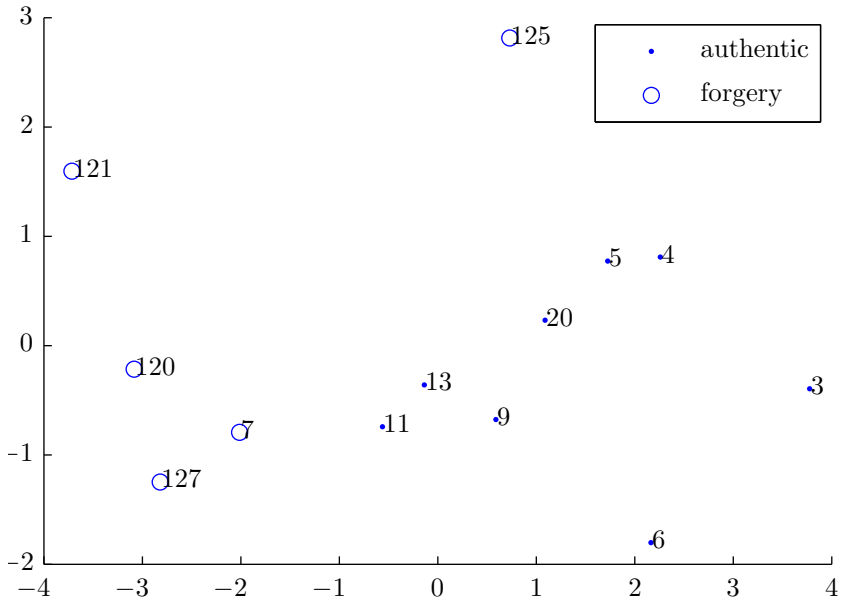


Figure 3.6: Visualization of the relationship between the training images; an embedding of points in $\mathbb{R}^2$ whose pairwise distances resemble those computed for the images.

To illustrate why the nearest neighbour classification fails to classify image 125 based on the remaining images, consider the embedding in Figure 3.7. The embedding is based on the weights computed using all images except 125 and the Voronoi tessellation [**7**] of the plane determined by the embedding. Image 125 is also included in the plot and here we see how the nearest neighbour classification determines the class of image 125.

The first thing we did was to test the influence of the EM algorithm: We trained 30 models from the original images and tested the classification scheme on them. For each of the models we computed the pairwise distances to the reference model. An embedding of points illustrating the pairwise distances is presented in Figure 3.8.

We wanted to test whether the distances from the trained model to the reference model are the same for each image. The distributions of these distances are not of the same shape and it is therefore not appropriate to test this hypothesis with an ANOVA or a non-parametric equivalent relying on assumptions about the shape of the distributions.

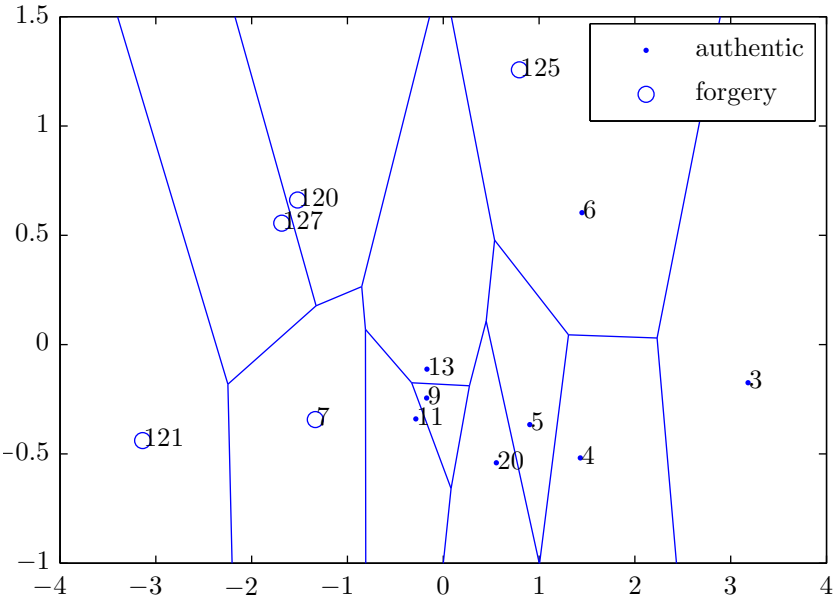


Figure 3.7: Illustration of how a nearest neighbour classification classifies image 125 based on the other images related to Pieter Bruegel; the lines show the Voronoi cells related to each point. Referring to Section 3.3 we have computed an optimal set of weights from all images except 125 using cross validation with nearest neighbour classification, and used these weights to compute pairwise distances between *all* images. From the pairwise distances we can classify image 125, and the embedding from multidimensional scaling shows a proxy of how this classification is performed.

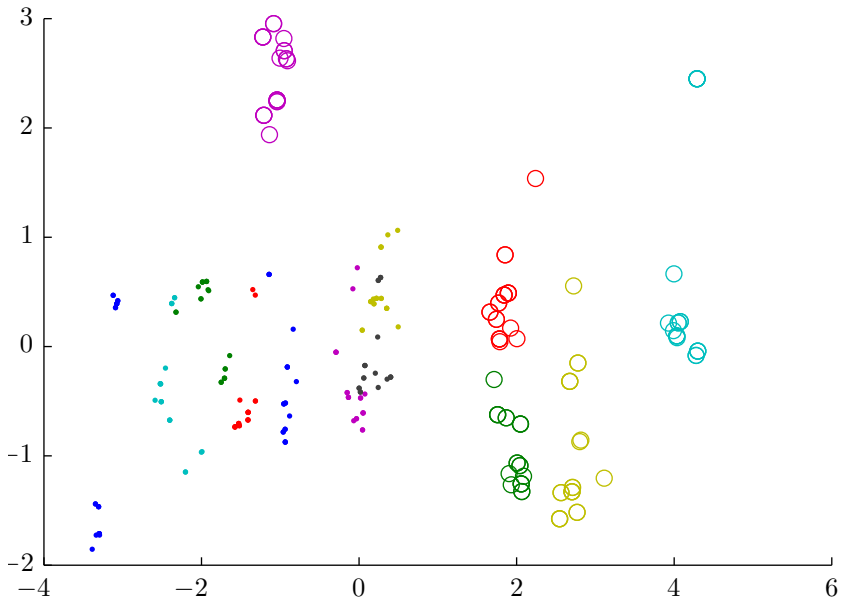


Figure 3.8: Visualization of the variation in the models introduced by the EM algorithm. With the weights computed from cross-validation of the reference models, the distance between each of the test models trained on the same image and random initialization of the EM algorithm were computed and embedded as described in Section 2.3.3. The models are coloured according to the image they represent; authentic and forgeries are represented using the same symbols as in Figure 3.6.

Instead we applied a permutation test [**11**] to test the hypothesis that the distribution of distances within each image are the same. If this hypothesis is indeed true, the distribution of distances will not change significantly under a random permutation of the labels. We performed a number of random permutations of the labels and calculated a test statistic for each of the permuted distances; this gave us a distribution of the test statistic where we can assess if the test statistic for the original configuration is likely. We had to use a test statistic that is reasonable for comparing the distances between the test images and the reference image and we have used the one-way ANOVA statistic.

In our case the hypothesis was rejected without doubt, as one would expect from the box plots of distances in Figure 3.9.

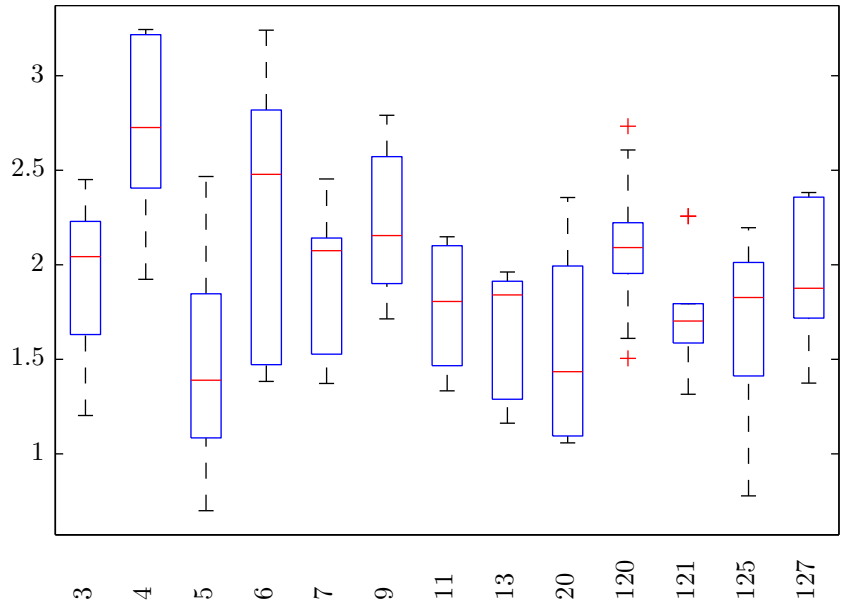


Figure 3.9: For each test model trained on the reference images we calculated the distance to the reference model shown in Figure 3.8. The box plots show the distribution of distances for each image. As we tested with a permutation test, these distributions are not alike.

Even though the EM algorithm does introduce variation, it is not so severe that it influences the classification; using a 1 nearest neighbour classification all of the test models were classified correct – as expected if the variations are reasonably small. As can be seen from Figure 3.8 using more than the nearest neighbour is not likely to correctly classify the image 125 – one point might have this as its nearest neighbour, but the second and third nearest point is likely to be an authentic image. On the other hand we might experience the opposite problem among the images that are closer (both authentic and forgeries); the closest neighbour might be of the wrong class, but the second and third closest neighbour are of the correct class.

3.4.1 Position experiments

Regarding the position experiments with the factors (1) and (2), we computed the classification error for each of the combinations and tested the variation of distances. The latter is performed individually for each image (due to the non-homogeneity of distances between the models of the reference images); we tested if the distances for a fixed value of (1) were of the same order of magnitude as for the reference models.

For the majority of the position experiments, the variation of distances are significantly higher than for the reference models. However, these factors have only little influence on the classification probabilities that are almost perfect with both 1 and 3 nearest neighbour classification – as seen in Table 3.2.

	nearest neighbours	
rotation	1	3
−4	0.97 ±0.04	0.91 ±0.032
−3	0.96 ±0.054	0.92 ±0
−2	0.98 ±0.032	0.92 ±0
−1	0.98 ±0.032	0.92 ±0
0	0.98 ±0.032	0.92 ±0.024
1	0.99 ±0.024	0.92 ±0.024
2	0.97 ±0.04	0.92 ±0
3	0.98 ±0.037	0.92 ±0
4	0.97 ±0.04	0.89 ±0.04

Table 3.2: Proportion of average correct classifications for the position experiments with k nearest neighbour classification. The classification rates for each rotation (specified by the degrees of rotation) is presented with average and standard deviation.

Performing the same experiments for the camera experiment, related to factors (3) to (6) we did not obtain as good classification results based on the reference model, but the results are interesting.

The number of correctly classified images varies between 5 and 12 out of the 13 possible with 1 nearest neighbour and between 8 and 12 for 3 nearest neighbours. As can be seen in Figure 3.10 the majority of the factor combinations have low classification rates.

Out of the 240 different combinations of factors roughly a quarter gets more than 8 correctly classified images with 3 nearest neighbours – and the common feature for all is that the contrast has not been increased.

After observing the effect from increasing the contrast, we examined the images with higher contrast and noticed that it disrupts the finest details – the very details that we are interested in modelling. We have included examples of this in Figure 3.11.

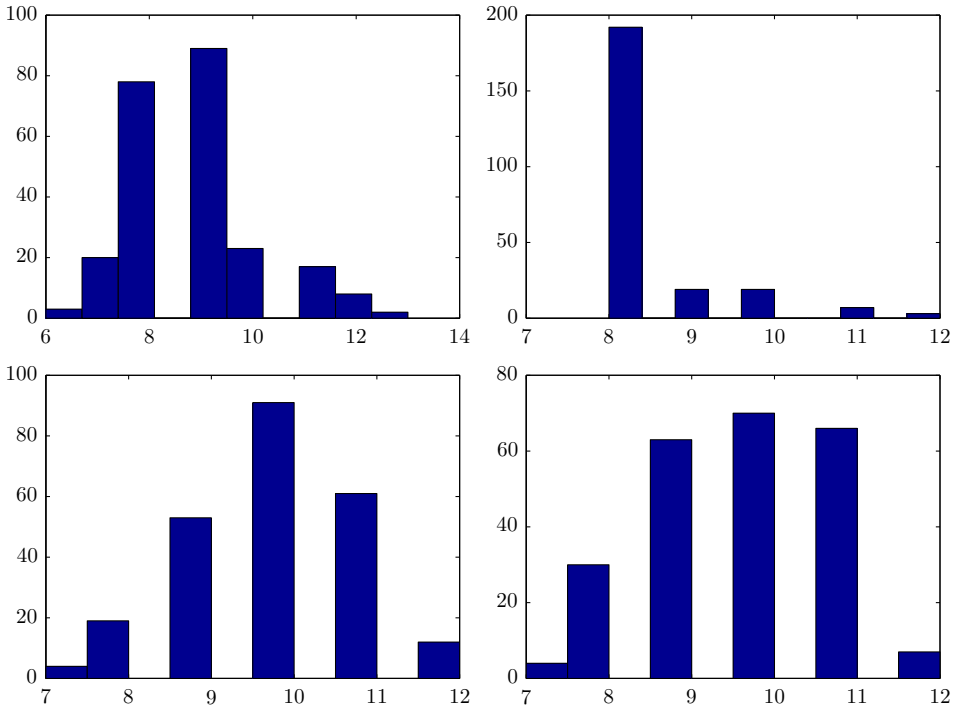


Figure 3.10: Histograms over the number of correctly classified images in the camera experiment. The top row is for classification with the reference model and the bottom row is for cross-validation. The left column is from classification using 1 nearest neighbour, while the right column is using 3 nearest neighbours.

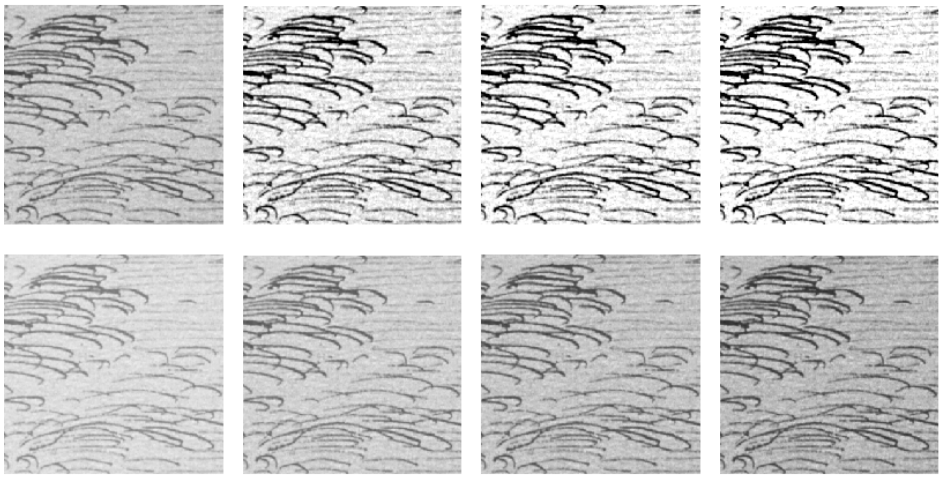


Figure 3.11: Examples of the effect of increasing contrast in a patch from the image 3. In the top row the contrast is increased by saturating as in (3.1) and from left to right 0, 2, 4 and 6% of the pixels have been saturated (with 0% being the original image). In the bottom row the contrast have been changed with (3.2) with γ having the values 0.5, 0.7, 0.9 and 1.1 from left to right. As can be seen this makes the light colors more difficult to distinguish in the patches to the left and vice versa with the rightmost patch.

The reason the brushstrokes gets corrupted by the contrast increasement is that the saturation (as explained in Section 3.3.2) introduces artifacts like enhancing the background so that it might look like brushstrokes.

Since our first approach to increasing contrast had such a substantial negative effect on the classification, we found it interesting to see if other contrast functions could improve the classification.

We therefore applied another contrast increasement function as explained in (3.2). The effect of this contrast increasement can also be seen in Figure 3.11.

The results from using this approach to increasing the contrast is presented in Table 3.3.

		γ of contrast curve											
k	kind	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2
1	class	5	5	5	7	10	10	13	13	13	13	13	13
1	cross	12	12	12	12	12	12	12	12	12	12	12	12
3	class	5	5	5	7	11	11	12	12	13	12	11	12
3	cross	11	11	11	12	12	12	12	12	12	12	12	12

Table 3.3: Number of correctly classified images with different contrast enhancements, parameterized by their γ value in (3.2). The number k is the number of nearest neighbours used in the classification. "kind" refers to the classification: "class" is classification by the reference models and "cross" is leave-one-out cross validation.

The high number of correct classifications in cross-validation indicates that even though the contrast changes, the images are still separable.

The reason why the number of correct classifications decrease so significantly for low values of γ is that these models look different from *all* the training models.

3.5 Conclusions

When applying a classification method to determine the authenticity of a painting, the method should not be too sensitive to small variations in the digital reproductions of the paintings. In this paper we simulated a number of factors that one is likely to encounter when acquiring digital reproductions and we tested their influence on our classification method from [**16**], which is based on second generation multiresolution analysis.

The factors related to the position of the camera relative to the paintings had no significant effect on our algorithm. The factors related to the recording properties of the camera may have some effect – especially the contrast of the image is influential if changed in the wrong way.

The experiments yielding negative effects of increasing the contrast by saturating extreme intensities led us to investigate if other contrast enhancements could have a

positive effect on the classification. Indeed, by increasing the contrast in a loss-less manner, our classification methods perform better.

We measured the influence of factor combination by performing leave-one-out cross-validation for the separability and classification with a reference model to see how the factors influence classification of "new" images. In both types of classification we classify up to all images correctly when avoiding the problems with change of contrast mentioned above.

Based on the experiments presented in this paper, we believe that our algorithms are suitable for real world data.

3.6 Acknowledgements

We are very grateful to Daniel Rockmore for providing us with the Bruegel images used in our experiments.

References

[1] Patrice Abry, Herwig Wendt, and Stéphane Jaffard. "When Van Gogh meets Mandelbrot: Multifractal classification of painting's texture". In: *Signal Processing* (2012). DOI: 10.1016/j.sigpro.2012.01.016.

[2] Mahmoud Al-Ayyoub, Mohammad T. Irfan, and David G. Stork. "Boosting multi-feature visual texture classifiers for the authentication of Jackson Pollock's drip paintings". In: *Proceedings of SPIE. Computer Vision and Image Analysis of Art II*. Ed. by David G. Stork, Jim Coddington, and Anna Bentkowska-Kafel. Vol. 7869. 2011.

[3] Igor Berezhnoy, Eric Postma, and H. Jaap van den Herik. "Authentic: Computerized Brushstroke Analysis". In: *IEEE International Conference on Multimedia and Expo*. 2005. DOI: 10.1109/ICME.2005.1521739.

[4] Igor E. Berezhnoy, Eric O. Postma, and H. Jaap van den Herik. "Automatic extraction of brushstroke orientation from paintings". In: *Machine Vision and Applications* 20.1 (Jan. 2009), pp. 1–9. DOI: 10.1007/s00138-007-0098-7.

[5] Duane C. Brown. "Decentering Distortion of Lenses". In: *Photometric Engineering* 32.3 (1966), pp. 444–462.

[6] Minh N. Do and Martin Vetterli. "The Contourlet Transform: An Efficient Directional Multiresolution Image Representation". In: *IEEE Transactions on Image Processing* 14.12 (Dec. 2005), pp. 2091–2106. DOI: 10.1109/TIP.2005.859376.

[7] Richard O. Duda, Peter E. Hart, and David G. Stork. *Pattern Classification*. 2nd ed. Wiley, 2001.

[8] David R. Falk, Dieter R. Brill, and David G. Stork. *Seeing the Light: Optics in Nature, Photography, Color, Vision, and Holography*. Harper & Row Publishers, New York, 1986.

[9] Rafael C. Gonzales and Richard E. Woods. *Digital Image Processing*. 2nd ed. Prentice Hall, 2002.

[10] Rafael C. Gonzales, Richard E. Woods, and Steven L. Eddins. *Digital Image Processing using Matlab*. Prentice Hall, 2004.

[11] Phillip Good. *Permutation Tests: A Pratical Guide to Resampling Methods for Testing Hypotheses*. Springer, 1994.

[12] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. 2nd ed. Springer, 2008. ISBN: 978-0-387-84857-0.

[13] Francis S. Hill. *Computer Graphics Using OpenGL*. 2nd ed. Prentice Hall, 2001.

[14] James M. Hughes, Daniel J. Graham, and Daniel N. Rockmore. "Quantification of artistic style through sparse coding analysis in the drawings of Pieter Bruegel the Elder". In: *Proceedings of the National Academy of Sciences* 107 (2009), pp. 1279–1283. DOI: 10.1073/pnas.0910530107.

[15] Mohammad Irfan and David G. Stork. "Multiple visual features for the computer authentication of Jackson Pollock's drip paintings: Beyond box-counting and fractals". In: *Proceedings of SPIE*. Image Processing: Machine Vision Applications II. Ed. by Kurt S. Niel and David Fofi. Vol. 7251. 2009. DOI: 10.1117/12.806245.

[16] C. Robert Jacobsen and Morten Nielsen. *Authentication of paintings using hidden Markov modelling of contourlet transforms*. Submitted.

[17] C. Richard Johnson, Jr et al. "Image Processing for Artist Identification". In: *IEEE Signal Processing Magazine* 25.4 (2008), pp. 37–48. DOI: 0.1109/MSP.2008.923513.

[18] Jia Li, Lei Yao, Ella Hendriks, and James Z. Wang. "Rhythmic Brushstrokes Distinguish van Gogh from His Contemporaries: Findings via Automated Brushstroke Extraction". In: *IEEE Transactions on Pattern Analysis & Applications* (Oct. 2011). DOI: 10.1109/TPAMI.2011.203.

[19] Siwei Lyu, Daniel Rockmore, and Hany Farid. "A digital technique for art authentication". In: *Proceedings of the National Academy of Sciences* 101.49 (Dec. 2004), pp. 17006–17010. DOI: 10.1073/pnas.0406398101.

[20] Laurens J. P. van der Maaten and Eric O. Postma. "Texton-based analysis of paintings". In: *Proceedings of SPIE*. Ed. by Andrew G. Tescher. Vol. 7798. Applications of Digital Image Processing XXXIII. 2010. DOI: 10.1117/12.863082.

[21] Duncan D.-Y. Po and Minh N. Do. "Directional Multiscale Modeling of Images using the Contourlet Transform". In: *IEEE Transactions on Image Processing* 15.6 (June 2006), pp. 1610–1620. DOI: 10.1109/TIP.2006.873450.

[22] Güngör Polatkan et al. "Detection of Forgery in Paintings Using Supervised Learning". In: *IEEE International Conference on Image Processing*. 2009. DOI: 10.1109/ICIP.2009.5413338.

[23] Jason P. de Villiers, F. Wilhelm Leuschner, and Ronelle Geldenhuys. "Centi-pixel accurate real-time inverse distortion correction". In: *Proceedings of SPIE*. Vol. 7266. 2008. DOI: 10.1117/12.804771.

[24] Bruce H. Walker. *Optical Engineering Fundamentals*. 2nd ed. SPIE Press, 2008.

CHAPTER 4

Comparing higher-order spatial statistics and perceptual judgements in the stylometric analysis of art

Publication details

Co-authors: James M. Hughes[†], Daniel J. Graham[+], Daniel N. Rockmore[‡†]

[†] *Department of Computer Science*
[‡] *Department of Mathematics*
[+] *Department of Psychological and Brain Sciences*
Dartmouth College

Journal: Proceedings of the European Signal Processing Conference 2011

Abstract:
Understanding the factors that underlie human perception of artistic style necessarily depends on measuring features beyond simple, low-order statistics. Indeed, much of our perception of style is rooted in qualities of lines and shading that cannot be described using first-order measures. In order to provide a richer, more accurate description of human perception of style, we must employ higher-order statistical methods. We demonstrate the applicability of two types of higher-order representations of images – and features derived from these – to the problem of similarity-based image search in large collections of art images. Our preliminary results indicate that a combination of perceptual information and statistical representations of art could prove extremely useful in navigating large art image databases in the context of similarity-based search.

4.1 Introduction

Increasingly, large collections of visual art are being digitized and made available through the World Wide Web (see e.g., www.googleartproject.com), both for enjoyment by the general public and for analysis by the scientific community. Of interest in such viewing experiences is the possibility of navigating these large image spaces according to concepts of visual similarity. This can be a challenging goal to accomplish, however, because of the complex statistical structure of visual art, which shares many basic statistical regularities with natural scenes. For example, spatial frequency amplitude spectra (second order statistics) in art images have roughly the same $1/f$-shaped falloff found in natural scenes [**9**, **10**, **22**]. A number of previous methods have shown promise in distinguishing and organizing artistic style (see e.g., [**15**]), but these techniques have typically not concentrated on describing works of art in terms of higher-order regularities.

Because low-level statistics are so widely shared across art types (even abstract works typically have very similar amplitude spectrum slopes to those found in representational works [**10**]), additional measures are needed to develop large-scale systems for predicting similarity among paintings and drawings in large collections of artworks. Such systems would be of potential use for the applications mentioned above, and we suggest that the principles guiding similarity judgment in artwork will go a long way towards improving content-based image retrieval systems more generally.

We argue that characterizing regularities in higher-order statistics presents a useful approach to the problem of organizing large art databases according to stylistic similarity (and, indeed, perceptual similarity). This suggestion is in line with past work showing that machine learning algorithms trained on higher-order spatial statistics are effective at performing fine-grained stylometric[1] distinctions [**13**]. It also aligns with neuroscientific evidence that sparse statistical regularities [**8**] shape neural coding strategies in the visual system [**6**, **7**]. That is, tailoring a "dictionary"

[1] "Stylometry" is a general term used to describe the development of quantitative tools for the analysis and understanding of artistic style.

of stylistic features to the higher-order redundancies found in art is akin to an organization schema of the receptive fields of visual neurons according to higher-order statistical regularities in nature [**19**].

Put another way, we efficiently adapt our representation to the higher order statistics of each image or image class, rather than using a standard orthonormal representation. This approach stands in contrast to the "kitchen sink" approach employed by other researchers [**23**], wherein one or more sets of features are chosen in an ad hoc way (e.g., RGB distributions, wavelet coefficients, face detection, etc.) to represent or describe a given image. While the latter approach has made important progress related to the analysis of large art databases – succeeding, for example, in separating art of different eras (e.g., Gothic vs. Impressionist) – there may be more principled ways to address the problems of quantifying style and using this information for image search and organization. Our approach, which includes making use of representations that are optimized for each image, attempts to provide a solution to these problems.

In addition, it remains to be seen whether style itself is a quality that is defined primarily with respect to perceived similarity as judged by lay viewers, historical or geographical provenance, or scholarly opinion. Psychological research in this vein – comparing style judgments in computer models trained on higher-order redundancies, art experts, and lay viewers – is underway. At present, though, there are but a handful of quantitative studies of the factors that govern human style perception.

Here we describe experiments demonstrating the effectiveness of representations that capture the higher-order statistics in a large, diverse collection of artworks. We compare this approach to other representational methods such as Gabor functions, as well as to approaches involving two-point statistics. We find that these statistical measures produce clusterings that agree in large part with a "true" underlying stylistic similarity between works of art, namely a labeling of the works based on the artist who created them. Further, we apply measures of higher-order redundancies to three sets of paintings of varying content (abstract art, landscapes, and portraits). A comparison of the resulting stylometric spaces to human judgements of similarity for the same image sets shows that approaches to stylometry using higher-order spatial regularities offer a promising route to capturing similarities across large, diverse collections of artworks.

We provide initial evidence that stylometric measures using higher-order statistical regularities show correspondences with perceived similarity. We believe that this suggests that systems that base judgements on human perceptual information, while at the same time taking advantage of as much quantitative information as possible (e.g., by including low-level statistics), are likely to provide the best performance in navigating image spaces using similarity-based search techniques.

4.2 Images

The images we used were a collection of 308 high-resolution art images obtain from various sources. Included are drawings by Bruegel [**21**], paintings by Charlotte Caspers [**3**], paintings by Georges Braque [**18**], drawings by Raymond Pettibon [**18**], and a large collection of works spanning several centuries obtained through the Cornell University collections [**9**, **10**], among others. All images were uncompressed TIFF or PNG images and were converted to grayscale via Matlab's `rgb2gray` function before analysis [**17**].

4.3 Image features

We examine the efficacy of two types of methods – fixed and adaptive – for providing descriptions of the stylistic qualities of art images. Furthermore, we compare these methods to the "expected" stylistic distinctions, as well as to psychophysical experiments that examined perceptual similarity between works of art [**11**]. The two image decomposition methods we utilize in this paper are a Gabor filter decomposition of images [**4**] and the sparse coding model [**19**, **20**]. Several features are extracted from the decompositions obtained using each of these models and are described in more detail in the corresponding sections.

Gabor filter decomposition

Gabor functions are localized, oriented, and bandpass, and as such are sensitive to constructs of lines and edges at particular orientations and spatial frequencies. In our experiments, we created a set of Gabor functions at eight orientations (0 to $\frac{7\pi}{8}$ radians), four spatial frequencies (approximately $5, 9, 12$, and 16 cycles-per-picture), and two phases (0 and π radians), for a total of 64 filters.

Once an image patch size (e.g., 64×64 pixels) and filter size (e.g., 32×32 pixels) were determined, we imposed a grid on the images and extracted as many patches of the specified size as possible. Each of these patches was convolved with the Gabor filters to generate a set of 64 filter responses. Generally, we let the filters have a side length equal to one-half the side length of the image patches. This allowed us to obtain a section of the convolution image equal in size to the filter, disregarding parts of the image where zero-padding would have been necessary.

Once the response images for each patch were obtained, a feature vector was generated for each patch using the energy contained in each filter response:

$$E(I, f_{k,\theta,\phi}) = \sum_i |(f_{k,\theta,\phi} * I)[i]|^2,$$

where I is the image patch and $f_{k,\theta,\phi}$ is a Gabor filter with preferred spatial frequency k, preferred orientation θ and spatial phase ϕ, and i indexes pixels in the image patch. Other features are of course possible, but for our purposes here we considered only this method of feature extraction. Distances between works of

art were determined by the correlation distance (i.e., $1 -$ Pearson's r) between the average of the feature vectors associated with a particular image.

Sparse coding model

The sparse coding model of Olshausen & Field [**19**, **20**], which is equivalent to independent component analysis (ICA) [**1**], was originally proposed to explain the response properties of cortical "simple cells" in the early visual system. The model learns a set of basis functions tuned to the higher-order statistical characteristics of a particular image space via maximum likelihood estimation. Since a sparse prior is used on the coefficients for any particular representation, the model attempts to maximize sparseness while guaranteeing a suitable level of reconstruction (i.e., one with relatively low reconstruction error).

For our purposes, we seek to take advantage of two important characteristics of this model: its sparseness, determined by non-Gaussian filter response distributions which allow the learned functions to be non-orthogonal and (possibly) overcomplete, and its adaptiveness, which insures that the learned functions are optimal with respect to the data. Sparseness is critical so that the functions do not become those that would be determined by a principal component analysis [**2**] decomposition of the image space, since such functions, which resemble the Fourier basis in two dimensions [**19**] and thus contain no localized information, are usually tuned to a narrow range of spatial frequencies and are generally not separable in terms of orientation and spatial frequency. Adaptiveness is also key: in contrast to a fixed decomposition such as a set of Gabor functions, the functions learned by the sparse coding model are data-dependent, and the variations in the properties of the functions themselves should be reflective of the underlying inputs.

Because of their adaptiveness to the input image space, we use the functions themselves as a proxy through which to analyze properties of the higher-order statistical characteristics of the images. Olshausen & Field showed that the learned functions reflect properties of the input image space [**19**]. We derive several features from the functions in order to analyze and compare these properties. In all of our experiments, we trained a set of 256 16×16 pixel basis functions on each image *individually* using the sparse coding model. It was from this set of functions that we derived features representing each image.

We compared images according to several metrics, which depend on the features extracted from the basis functions. They are as follows:

- Peak orientation: given the two-dimensional Fourier transform $F(\omega, \theta)$ of a basis function (viewed as a function of frequency ω and angle θ), we find the orientation θ^* at which peak amplitude (or power) occurs, averaged across all spatial frequencies, i.e.,

$$\theta^* = \arg\max_{\theta} \frac{1}{|\Omega|} \sum_{\omega \in \Omega} |F(\omega, \theta)|.$$

  This is a reliable way of determining the orientation selectivity of a basis function.

- Peak spatial frequency: given the two-dimensional Fourier transform $F(\omega, \theta)$ of a basis function, we find the spatial frequency ω^* at which peak amplitude (or power) occurs, averaged across all orientations, i.e.,

$$\omega^* = \arg\max_{\omega} \frac{1}{|\Theta|} \sum_{\theta \in \Theta} |F(\omega, \theta)|.$$

  This is a reliable way of determining the spatial frequency selectivity of a basis function.

- Orientation bandwidth: given the two-dimensional Fourier transform of a basis function, we find the bandwidth in octaves (measured by full width at half-maximum) of the function, averaged across all spatial frequencies, centered around its peak orientation θ^* (as given above). This quantity measures how selective a basis function is for its preferred orientation.

- Spatial frequency bandwidth: given the two-dimensional Fourier transform of a basis function, what is the bandwidth in octaves (measured by full width at half-maximum) of the function, averaged across all orientations, centered around its peak spatial frequency ω^* (as given above). This quantity measures how selective a basis function is for its preferred spatial frequency.

These quantities are computed for each of the 256 basis functions trained for each image. Since there is no natural way to compare individual functions with one another, we employ distributional methods to do so. In particular, we use symmetrized Kullback-Leibler divergence (KLD) to compare distributions of these quantities, defined in the following way:

$$\mathrm{KLD}(P, Q) = \frac{1}{2} \sum_{\omega \in \Omega} \left[P(\omega) \log \frac{P(\omega)}{Q(\omega)} + Q(\omega) \log \frac{Q(\omega)}{P(\omega)} \right].$$

Since the values above are continuous quantities, we estimate KLD by binning, and we determine bins *once* for a particular quantity (e.g., spatial frequency bandwidth), and this determines the binning for all subsequent computations of the KLD. Thus, given the quantities above, we derive distances between all images using KLD for the following distributions:

- Distribution of peak orientation
- Distribution of peak spatial frequency
- Joint distribution of peak orientation and spatial frequency
- Distribution of orientation bandwidth
- Distribution of spatial frequency bandwidth
- Joint distribution of orientation and spatial frequency bandwidth

Furthermore, we compute distances between images based on a distance metric defined directly on the sets of basis functions [**14**]. The final feature we compute, from which we derive a distance, is the slope of the log rotational average of the amplitude spectrum for each image [**6**]. Ultimately, including Gabor filter energy, all of the distances derived from the sparse coding model basis functions, and the slope of the log rotational average of the amplitude spectrum, we have in total ten distances with which we compare images in our dataset. We include an eleventh, the distance matrix derived by aggregating the distance matrices after rescaling each so that the maximum distance was 1.

4.4 Methods

Our ultimate goal in this work is to provide a method for quantitatively characterizing the style of a work of art, specifically in such a way that allows us to easily compare a given work with others and determine which are stylistically most similar. Unfortunately, we do not at present possess a "ground truth" notion of stylistic similarity for our entire dataset. In order to compare the various derived metrics on images, we chose to compare them according to the true artist labeling of the images. For example, all paintings by Picasso would have the same label. This rule was used consistently except in one prominent case, the "non-Bruegel" category of drawings, in which drawings were given the same label, though they may be by different artists. This oversmoothing was necessary since the attribution of many of the Bruegel imitation images is not known; however, they are, like the Bruegel drawings, fairly stylistically consistent, especially with respect to the other works of art in the dataset.

In order to compare the information contained in the ten distance matrices with the true artist labeling of the images, we first embedded the drawings in Euclidean space via classical multidimensional scaling [**12**]. We then used the k-means algorithm to determine a clustering of the points for several values of k (shown in Figure 4.1). Once a clustering was determined for the embedding of each distance matrix, we compared that clustering with the true labeling using normalized mutual information (NMI) [**5**, **16**]:

$$\mathrm{NMI}(\Omega, C) = \frac{I(\Omega, C)}{[H(\Omega) + H(C)]/2},$$

where I is the mutual information between clustering Ω and the true labeling C,

$$I(\Omega, C) = \sum_k \sum_j P(\omega_k \cap c_j) \log \frac{P(\omega_k \cap c_j)}{P(\omega_k) P(c_j)},$$

and H is the entropy of each set of objects,

$$H(\Omega) = -\sum_k P(\omega_k) \log P(\omega_k).$$

The NMI between two clusterings is a number in the unit interval that reflects their information overlap. NMI is not the only measure for information overlap,

but it serves as a sound proxy for our purposes in these initial experiments, due to two desirable properties, namely that: 1) it is a bounded measure that favors good overlaps, and 2) it penalizes over fitting, i.e. using too many classes.

4.5 Results

Figure 4.1 shows the value of NMI for clusterings obtained via k-means clustering using each of the distance metrics described above, for several values of k. As can be seen, the best performance was obtained using the combined (i.e., aggregated) distance matrix, with the Euclidean distance-based basis metric, Gabor filter energies, and joint orientation/spatial frequency bandwidth distributions also containing information that was consistent with the true labeling of the images. The overall success of the combined distance matrix suggests that incorporating statistical measures based on fixed representations (such as Gabor filter energies and the slope of the amplitude spectrum) as well as adaptive measures (like those that depend on basis functions learned via a sparse coding model) is the most effective way to characterize stylistic properties in works of art.

Comparison to perceptual experiments

In order to begin to evaluate the effectiveness of the derived features at accounting for *perceptual* similarities between images, we compared our feature-based distances to those obtained through psychophysical perceptual similarity experiments [**11**]. In these experiments, participants were asked to judge the similarity between pairs of art images (1-9 scale) in three categories, abstract art, landscapes, and portraits (images were sorted in a prior three-alternative forced choice test by a separate set of subjects). This information was aggregated across all participants to create a similarity matrix for each category. These experiments were extremely small scale and dealt with approximately 20 images per category; nevertheless, as we will demonstrate, the perceptual similarities between works of art provide information that is effective at categorizing images according to their style, at least at the coarse artist-by-artist scale of our experiments.

The first experiment we performed was to compare the effectiveness of the perceptual judgements at predicting the stylistic relationship between works of art. Given a set of images I, we have perceptual distances $D_{jk} = D(I_j, I_k)$ between all pairs of images. We trained a regression model on the distances D_{jk} using the feature-based distances as regressors (in our model, we include constant, linear, and quadratic terms). In this fashion, we should be able to predict the relationship d_{jk} between any two randomly held-out images, according to the perceived similarity between the remaining images, assuming the perceptual distances contain useful information that will be predictive for the held-out images. We performed the following experiment 500 times:

- Hold out two images at random, say I_{j_1}, I_{j_2}

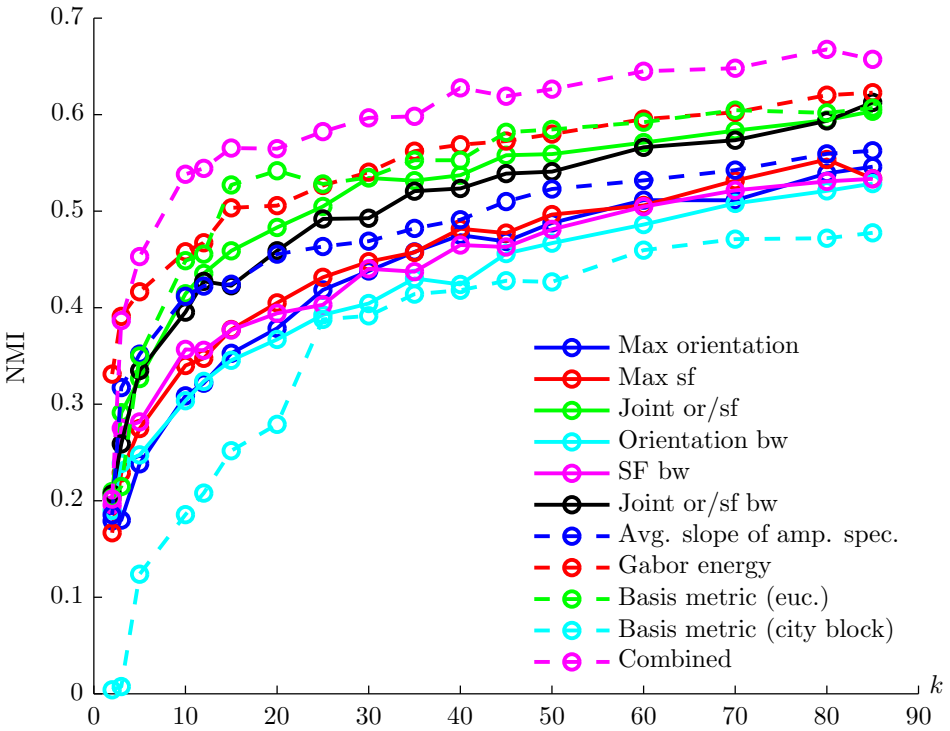


Figure 4.1: Clustering normalized mutual information for the 308 images used in our experiments, clustered using the k-means algorithm, across several values of k (indicated by open circles on traces in plot). A maximum of 85 clusters was chosen, since that corresponds to the number of unique artist labels in the dataset. A perfect clustering would have NMI of 1.

- Calculate a regression model on all images $I \setminus \{I_{j_1}, I_{j_2}\}$
- Predict the distances between the held-out images and the training images, i.e., $d(I_{j_1}, I_p), d(I_{j_2}, I_p)$, for all $p \neq j_1, j_2$
- Compare the predicted distances d_{jk} with the true perceptual distances for one image I_p selected at random; if $D(I_{j_1}, I_p) > D(I_{j_2}, I_p)$ and $d(I_{j_1}, I_p) > d(I_{j_2}, I_p)$, then the prediction is deemed correct.

Accuracy and statistical significance of the results of these experiments for each of the three classes is shown in Table 4.1.

Category	Accuracy	p-value
Abstract art	0.61	5×10^{-7}
Landscapes	0.62	1×10^{-8}
Portraits	0.51	0.28

Table 4.1: Accuracy of model at predicting relationship of held-out images to randomly selected image in training set. High accuracy implies that the correct relationship between the test images was predicted by the model. Accuracy is given by the fraction of correct predictions (out of 500 tests). The right column shows p-values indicating the significance level of predicting the corresponding number of correct similarity relationships, assuming a binomial model in which the correct relationship would be guessed at random (i.e., by flipping a fair coin, where "heads" indicates that the correct relationship was observed).

These results indicate that perceptual information from two of the three categories (abstract art and landscapes) contained information that allowed prediction at a statistically significant level. This not only confirms the existence of useful information in the perceptual similarity data, but also the ability of the statistical information to effectively model these distinctions. Nevertheless, the number of images used in these experiments was extremely limited, and in order to generalize these results, further experiments are required.

We also explored the extent to which the admittedly limited perceptual information we possessed about the three categories of images could be predictive of stylistic distinctions in larger sets of images. This application is of particular importance to similarity-based image search, since its success implies that limited subsets of perceptual similarity information between images could be used to "bootstrap" models used to predict similarity between very large sets of images.

Prediction of perceptual distances between images in our dataset was accomplished by first modeling the perceptual distances in the three image categories using the feature-based scores, as before, but this time without holding out any images. Once a model was obtained, we averaged the the predictor coefficients from each of the three models and used this to predict the "perceptual distance" between images. Note that, in these experiments, we did not predict distances for the

images that were included in the original perceptual experiments (i.e., those in the abstract, landscape, and portrait categories). These images were held-out during the prediction phase and did not factor into the subsequent analysis.

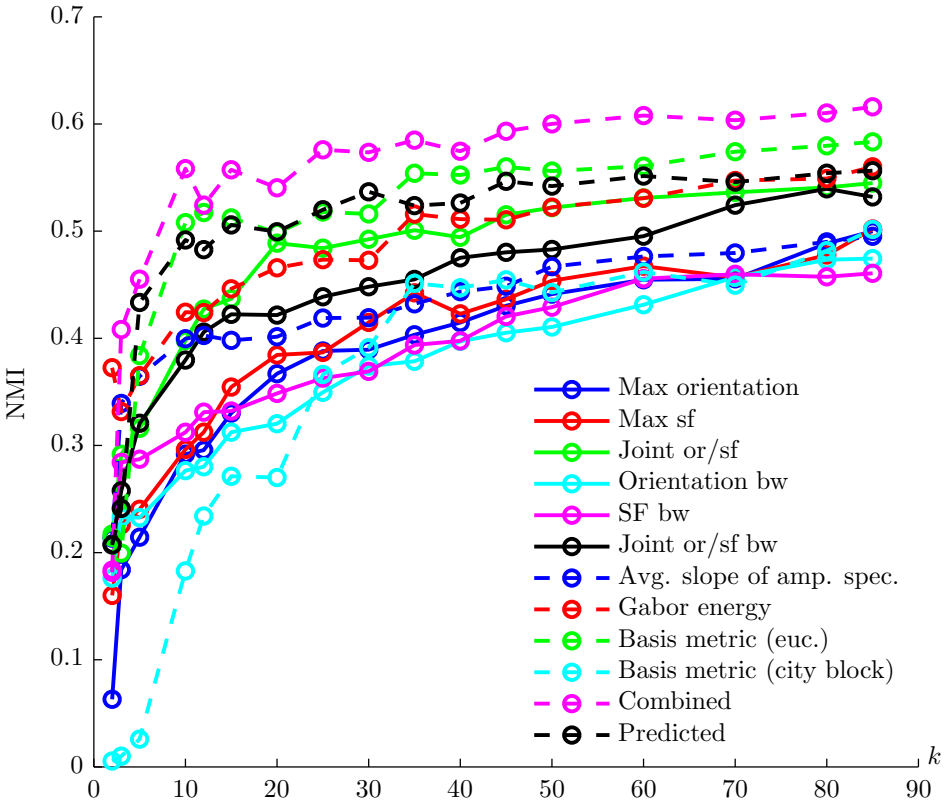


Figure 4.2: Clustering normalized mutual information for 257 images used to compare clusterings between distance measures and predicted perceptual distances, clustered using the k-means algorithm, across several values of k (indicated by open circles on traces in plot). A perfect clustering would have NMI of 1.

We held out 51 images (across the three categories) and created a predicted perceptual distance matrix on the remaining 257 images. We compared this with the same distance measures used above by using only the submatrices that contained these 257 works in each of the 11 original matrices. As before, we found a Euclidean embedding of the images using each distance matrix and then performed k-means clustering for several values of k, then compared these clusterings with the corresponding true labeling. The results are shown in Figure 4.2. Not surprisingly, the aggregate distance matrix still yields clusterings with the closest relationship to the true labeling. However, the predicted distance matrix has performance on par with any other individual feature, suggesting that, although our sample of perceptual distances was extremely limited, this information can guide the ways in which we

combine statistical features to understand style perception.

4.6 Conclusions

Although preliminary, our results indicate not only that measurement of higher-order statistical characteristics of images generates information germane to stylistic distinctions, but also that combining this information with perceptual similarity can create an effective, statistical means of organizing images and predicting perceptual similarity in the context of similarity-based search in large image databases. Future work will include implementation of such a system and further analysis of the concepts presented here.

4.7 Acknowledgements

JMH, DJG, and DNR gratefully acknowledge the partial support of the Samuel H. Kress Foundation for this work.

References

[1] Anthony J. Bell and Terrence J. Sejnowski. "An information-maximization approach to blind separation and blind deconvolution". In: *Neural Computation* 7.6 (1995), pp. 1129–1159. DOI: 10.1162/neco.1995.7.6.1129.

[2] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2007. ISBN: 978-0-387-31073-2.

[3] *Charlotte data set*. URL: http://www.math.princeton.edu/ipai/datasets.html.

[4] John G. Daugman. "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters". In: *Journal of the Optical Society of America A* 2.7 (1985), pp. 1160–1169. DOI: 10.1364/JOSAA.2.001160.

[5] Richard O. Duda, Peter E. Hart, and David G. Stork. *Pattern Classification*. 2nd ed. Wiley, 2001.

[6] David J. Field. "Relations between the statistics of natural images and the response profiles of cortical cells". In: *Journal of the Optical Society of America A* 4 (1987), pp. 2379–2394. DOI: 10.1364/JOSAA.4.002379.

[7] David J. Field. "What is the goal of sensory coding?" In: *Neural Computation* 6.4 (1994), pp. 559–601. DOI: 10.1162/neco.1994.6.4.559.

[8] Daniel J. Graham and David J. Field. "Sparse coding in the neocortex". In: *Evolution of Nervous Systems* 3 (2007). Ed. by John H. Kaas and Leah A. Krubitzer, pp. 181–187. DOI: 10.1016/B0-12-370878-8/00064-1.

[9] Daniel J. Graham and David J. Field. "Statistical regularities of art images and natural scenes: spectra, sparseness and nonlinearities". In: *Spatial Vision* 21.1 (2007), pp. 149–164. DOI: 10.1163/156856808782713771.

[10] Daniel J. Graham and David J. Field. "Variations in intensity statistics for representational and abstract art, and for art from the eastern and western hemispheres". In: *Perception* 37.9 (2008), pp. 1341–1352. DOI: 10.1068/p5971.

[11] Daniel J. Graham, J. D. Friedenberg, Daniel N. Rockmore, and David J. Field. "Mapping the similarity space of paintings: image statistics and visual perception". In: *Visual Cognition* 18.4 (2010), pp. 559–573. DOI: 10.1080/13506280902934454.

[12] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. 2nd ed. Springer, 2008. ISBN: 978-0-387-84857-0.

[13] James M. Hughes, Daniel J. Graham, and Daniel N. Rockmore. "Quantification of artistic style through sparse coding analysis in the drawings of Pieter Bruegel the Elder". In: *Proceedings of the National Academy of Sciences* 107 (2009), pp. 1279–1283. DOI: 10.1073/pnas.0910530107.

[14] C. Robert Jacobsen, James M. Hughes, Daniel J. Graham, and Daniel N. Rockmore. "A distance metric between sets of vectors". In preparation.

[15] C. Richard Johnson, Jr. et al. "Image Processing for Artist Identification". In: *IEEE Signal Processing Magazine* 25.4 (2008), pp. 37–48. DOI: 0.1109/MSP.2008.923513.

[16] Christopher Manning, Prabhakar Raghavan, and Hinrich Schütze. *An Introduction to Information Retrieval*. Cambridge University Press, 2009. ISBN: 9780521865715.

[17] *Matlab software*. The Mathworks, Natick, MA. 2011.

[18] *Museum of Modern Art, New York City*. Courtesy Jim Coddington.

[19] Bruno A. Olshausen and David J. Field. "Emergence of Simple-Cell Receptive Field Properties by Learning a Sparse Code for Natural Images". In: *Nature* 381 (June 1996), pp. 607–609. DOI: 10.1038/381607a0.

[20] Bruno A. Olshausen and David J. Field. "Sparse Coding with an Overcomplete Basis Set: A Strategy Employed by V1?" In: *Vision Research* 37 (Dec. 1997), pp. 3311–3325. DOI: 10.1016/S0042-6989(97)00169-7.

[21] Nadine M. Orenstein, ed. *Pieter Bruegel the Elder – Drawings and Prints*. Metropolitan Museum of Art, New York City and Yale University Press, New Haven, 2001.

[22] Christoph Redies, Jens Hasenstein, and Joachim Denzler. "Fractal-like image statistics in visual art: similarity to natural scenes". In: *Spatial Vision* 21.1-2 (2007), pp. 137–148. DOI: 10.1163/156856808782713825.

[23] Christian Wallraven et al. "Categorizing art: Comparing humans and computers". In: *Computers and Graphics* 33.4 (2009), pp. 484–495. DOI: 10.1016/j.cag.2009.04.003.

CHAPTER 5

Unpublished work

5.1 Curve metric

The work presented in this section is an attempt to find a meaningful metric between one-dimensional signals.

The ideas presented in the following chapter are motivated by considering the data in Figure 5.1 and the reader can refer to this figure for visualization. The data in Figure 5.1 were originally presented in [**64**] and was provided for this study by courtesy of Mark de Zee, Department of Health Science and Technology and John Rasmussen, Department of Mechanical and Manufacturing Engineering, both at Aalborg University. The following description of the experiment and nature of the data is from personal communication with Mark de Zee.

The signals were collected to validate a musculo-skeletal model for the human mandible by measuring EMG activity of the muscles being modelled. Over a period of 10 seconds participant were asked to tense their muscles gradually and then relax them. Afterwards the measurements were normalized with respect to the maximum measured EMG value for the given muscle; the maximum value was determined by asking the participant to tense the muscle as much as possible.

The signals occur in pairs; one is the measured electrical impulse from the test subject and the other is the impulse computed by the model. If the model is accurate the two signals should look similar – the matter of defining similarity is the issue here.

Many of the signals in Figure 5.1 seem to have the same form – except that the measured signal is more irregular than the simulated signal, as is the case in Figure 5.1a. One request to a metric between such signals is that signal pairs of this

form should be "close", whereas the signals in e.g. Figure 5.1f should be further apart.

Simply using an ℓ^p-norm as a metric between curves is not a good idea for at least two reasons:

- The ℓ^p-norm is too sensitive to high frequency content.
- If the signals change rapidly over a small time frame, the ℓ^p-norm can change drastically. This situation can for instance occur if the signal measurement are delayed.

The basic idea for a curve metric fulfilling these demands is based on filtering. Consider the case in e.g. Figure 5.1b where the two curves have pretty much the same shape, except that the measured signal have more high frequency fluctuations. Another way to phrase this observation is that if we apply a low pass filtering on the observed signal, they would be more alike. In this situation one could be inclined to ask how much filtering is appropriate – and this could very well be highly dependent on the actual signals. A major point of having a metric between such curves is that we should not have to tune parameters like the amount of filtering, so this is not an attractive approach.

However, if we apply the same filtering to both curves in consideration we have two low pass approximation that I believe will provide us with new insight. If a curve has less high frequency information, the low pass filtering of this curve will not remove much information. But if the curve has many small fluctuations, the low pass filtering will give us a significantly smoothed version.

So in the situation where two curves are alike apart from high frequency content (in one or both), a low pass filtering of both curves will leave them more alike.

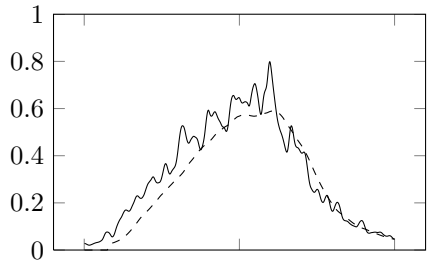
This argument can be repeated, so that a coarser low pass filtering/successive low pass filterings will give us curves that should be even more alike.

Instead of choosing *the* right amount of filtering for the comparison, I propose to define a metric between signals by comparing successive low pass filterings of their difference by an ℓ^2-norm.

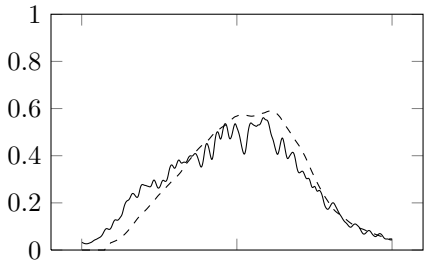
Let us now formalize this. Besides fulfilling the definitions for being a (pseudo-)metric, it is natural to demand that a metric is suited for the problem at hand. In the context of a metric between curves at least two properties are reasonable to demand:

- Scale invariance. The distance between curves should not depend on the measurement scale used.
- Good interaction with translation operators. To circumvent the problem described in the beginning of the section with too large distances between signals being shifted relative to each other.

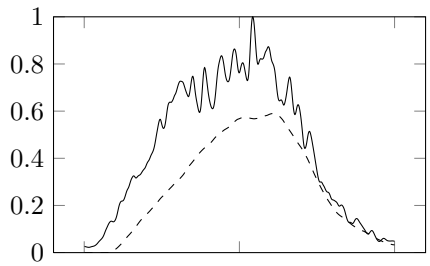
In the current proposal for a metric the first issue is not being considered, since the data we are working with has been normalized. Incorporating scale independence



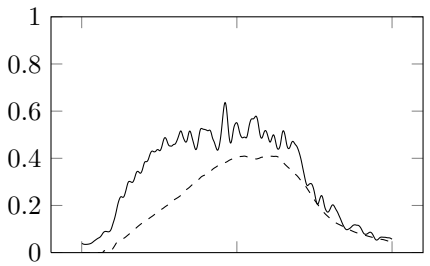
(a) Masseter left, distance is 1.16.



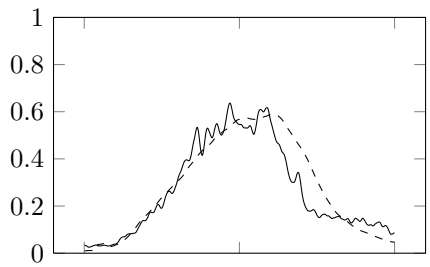
(b) Masseter right, distance is 1.12.



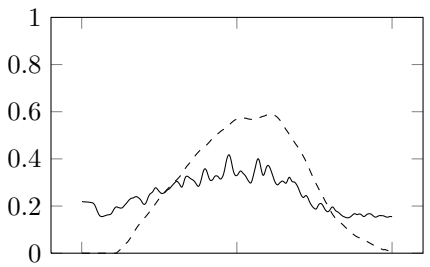
(c) Temporalis left, distance is 3.33.



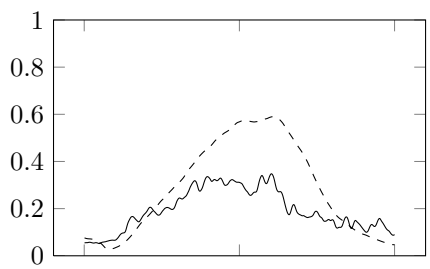
(d) Temporalis right, distance is 2.58.



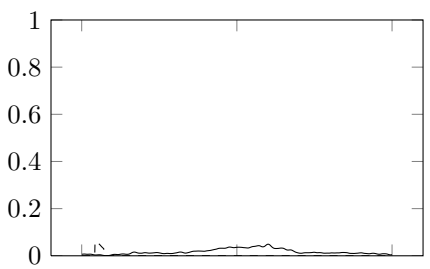
(e) Medial pterygoid left, distance is 0.65.



(f) Medial pterygoid right, distance is 2.54.



(g) Superior lateral pterygoid left, distance is 2.05.



(h) Superior lateral pterygoid right, distance is 0.29.

Figure 5.1: Data used for testing the curve metric. In each of the sub figures the solid curve is the measured signal and the dashed curve is the computed signal. The distance calculated with the curve metric (using a Daubechies 4 filter) between the two signals in each sub figure is written below the figure along with the muscle involved in the experiment.

is not a trivial task and should be considered if the metric is going to be used for more general data, but for the data at hand this is not a problem.

As a matter fact using metrics that does not take the normalization into account and therefore includes some kind of normalization can lead to strange results. For the data tested here a metric designed for comparing waveforms [**27**] have been tried [**44**], but especially the distance between the signals in Figure 5.1h are counterintuitive because they are deemed very different (John Rasmussen, Aalborg University, personal communication).

The issue of translation dependence will be dealt with below.

The filtration is performed with a wavelet transform and in order to deal with the translation dependence we use a stationary wavelet transform, see e.g. [**47**].

When obtaining a multiresolution analysis of a discrete signal from the discrete wavelet transform, we apply the same low pass and high pass filter in each level of the transform. Furthermore, we downsample the low pass part after applying a filter, thereby discarding every other observation. The downsampling is what causes the troublesome interaction with translations for the ordinary wavelet transform.

The stationary wavelet transform is obtained from the usual discrete wavelet transform by only a slight modification: In the stationary wavelet transform we do not downsample – instead we modify the filters in each level of the multiresolution analysis by upsampling.

For the remainder of this section we use the following notation:

- $\boldsymbol{x} = \{x[i]\}_{i\in\mathbb{Z}}$ a discrete, one dimensional signal in $\ell^2(\mathbb{Z})$.
- $\mathcal{F} : \ell^2(\mathbb{Z}) \to L^2([-\pi, \pi])$ is the Fourier operator.
- $L_k : \ell^2(\mathbb{Z}) \to \ell^2(\mathbb{Z})$ is a low pass operator. In words L_k removes the high frequency content of a signal. In mathematical terms we can characterize L_k by its behavior in the Fourier domain:

$$(\mathcal{F}L_k\boldsymbol{x})(\xi) = m_k(\xi)(\mathcal{F}\boldsymbol{x})(\xi), \tag{5.1}$$

  where the $m_k : [-\pi, \pi] \to [0, 1]$ are piece-wise smooth cut-off functions: They are 1 around the origin, 0 away from the origin and the supports are strictly decreasing as k grows.

- $\boldsymbol{x}_k$ is the k'th low pass filtering of $\boldsymbol{x}_k = L_k\boldsymbol{x}$.
- $\tau_h : \ell^2(\mathbb{Z}) \to \ell^2(\mathbb{Z})$ is translation by $h \in \mathbb{Z}$, $(\tau_h\boldsymbol{x})[i] = x[i - h]$. Due to the representation of a signals index by integers, we can only translate by integers as well. But, as in all sampling, one must bear in mind that the integers represent a choice of scale and therefore integer translations are not necessarily "large".

A low pass filtering of $\boldsymbol{x}$ is obtained by decomposing a signal into a low frequency and high frequency part with the stationary wavelet transform, discarding the high frequency part and then reconstructing the signal with the inverse stationary wavelet transform using only the low frequency part.

Due to the lack of downsampling in the stationary wavelet transform, the associated low pass filter commutes with the translation operator:

$$L(\tau_h \boldsymbol{x}) = \tau_h(L\boldsymbol{x}).$$

When applying the wavelet transform to a finite signal we need to perform a suitable correction at the boundaries by extending the signal. The best choice of boundary correction is generally problem dependent and in the experiments with the data in Figure 5.1 I have used zero padding. The reason for using this type of extension is because of the nature of the signals: They were measured over a fixed time interval, so measurements beyond this time frame would be discarded and the test persons are assumed to relax their muscles in the beginning and end of the experiments, thereby yielding small EMG values.

As mentioned in the motivation, the idea behind the metric soon to be defined is to perform successive filtrations, so we must also choose the number of filtrations M. For a finite signal of length $N = 2^M$, we can successively apply a low pass filtration M times.

Let $\{a_k\}_{k=1}^M$ be an increasing sequence of positive numbers, such that

$$\sum_{k=1}^{M} a_k \leq 1. \tag{5.2}$$

In the experiments performed I have used $a_k = 2^{M-k}$.

The proposed metric between two signals $\boldsymbol{x}$ and $\boldsymbol{y}$ is defined as

$$d(\boldsymbol{x}, \boldsymbol{y}) := \sum_{k=1}^{M} a_k \|\boldsymbol{x}_k - \boldsymbol{y}_k\|_{\ell^2(\mathbb{Z})}. \tag{5.3}$$

It should be noticed that since we compare only filtered versions of the signals, (5.3) is not faithful, i.e., two non-identical signals can have distance zero. I would argue, however, that this is desirable in the current situation since two signals whose low pass filterings are identical must differ only by high frequency content (i.e., noise) and for practical issues this means that they are identical.

From the properties of the norm it is seen that (5.3) is symmetric and obeys the triangle inequality, and therefore is a pseudo-metric (or, in fact, a semi-norm).

The low pass filters from (5.1) satisfies that $|m_k| \leq 1$ and with Parseval's identity we get that

$$\begin{aligned} \|\boldsymbol{x}_k - \boldsymbol{y}_k\|_{\ell^2(\mathbb{Z})} &= \frac{1}{2\pi}\|\mathcal{F}(\boldsymbol{x}_k - \boldsymbol{y}_k)\|_{L^2([-\pi,\pi])} = \frac{1}{2\pi}\|m_k \mathcal{F}(\boldsymbol{x} - \boldsymbol{y})\|_{L^2([-\pi,\pi])} \\ &\leq \frac{1}{2\pi}\|\mathcal{F}(\boldsymbol{x} - \boldsymbol{y})\|_{L^2([-\pi,\pi])} = \|\boldsymbol{x} - \boldsymbol{y}\|_{\ell^2(\mathbb{Z})}. \end{aligned} \tag{5.4}$$

So with the requirement (5.2) we have that the metric fulfills

$$d(\boldsymbol{x}, \boldsymbol{y}) = \sum_{k=1}^{M} a_k \|\boldsymbol{x}_k - \boldsymbol{y}_k\|_{\ell^2(\mathbb{Z})} \leq \|\boldsymbol{x} - \boldsymbol{y}\|_{\ell^2(\mathbb{Z})} \sum_{k=1}^{M} a_k \leq \|\boldsymbol{x} - \boldsymbol{y}\|_{\ell^2(\mathbb{Z})}.$$

To elaborate on how this metric interacts with translation operators, we need the following elementary property of the Fourier transform, connecting translations and modulations:

$$\big(\mathcal{F}(\tau_h \boldsymbol{x})\big)(\xi) = e^{-i\xi h}(\mathcal{F}\boldsymbol{x})(\xi). \tag{5.5}$$

Furthermore, we need to bound the complex exponential function in the following manner:

$$|1 - e^{it}| \leq 2 \wedge |t|, \quad t \in \mathbb{R}. \tag{5.6}$$

With these properties and Parseval's identity we can obtain the following bound on the L^2-norm of the difference between a low pass filtering $\boldsymbol{x}_k$ and a translated version of the same signal:

$$\begin{aligned}
\|\boldsymbol{x}_k - \tau_h \boldsymbol{x}_k\|_{\ell^2(\mathbb{Z})} &= \frac{1}{2\pi}\|\mathcal{F}\big(\boldsymbol{x}_k - \tau_h \boldsymbol{x}_k\big)\|_{L^2([-\pi,\pi])} \\
&= \frac{1}{2\pi}\|(\mathcal{F}\boldsymbol{x}_k)(\xi)\big(1 - e^{-i\xi h}\big)\|_{L^2([-\pi,\pi])} \\
&\leq \frac{1}{2\pi}\|(\mathcal{F}\boldsymbol{x}_k)(\xi)\big(2 \wedge |\xi h|\big)\|_{L^2([-\pi,\pi])}
\end{aligned}$$

To provide further bounds on this quantity, the nature of functions in the norm should be remembered.

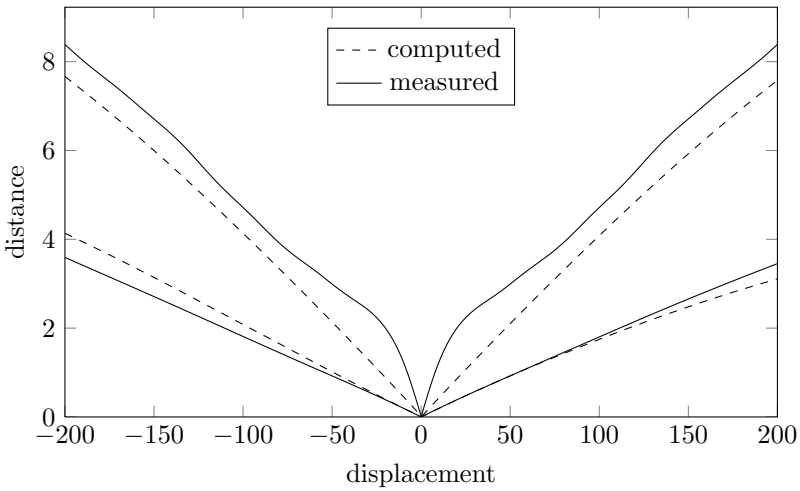
The inequality (5.6) is only of interest for numerically small values of t and due to the requirement that $h \in \mathbb{Z}$, the minimum $2 \wedge |\xi h|$ would quickly favor 2 if we consider a general $\xi \in [-\pi, \pi]$. However, just as in (5.4) we have that the support of $\mathcal{F}\boldsymbol{x}_k$ is centered around the origin – and that it can be much smaller than that of $\mathcal{F}\boldsymbol{x}$. Therefore, especially for the coarse filtrations and reasonably small translations, $|\xi h| \leq 2$ – and in any case the bound $2 \wedge |\xi h| \leq |\xi h|$ holds. So if we continue the calculations above we have that

$$\begin{aligned}
\|\boldsymbol{x}_k - \tau_h \boldsymbol{x}_k\|_{\ell^2(\mathbb{Z})} &\leq |h|\frac{1}{2\pi}\|\xi(\mathcal{F}\boldsymbol{x}_k)(\xi)\|_{L^2([-\pi,\pi])} \\
&\leq \frac{|h|}{2}\|\mathcal{F}\boldsymbol{x}_k\|_{L^2([-\pi,\pi])}) = \frac{h}{2}\|\boldsymbol{x}_k\|_{\ell^2(\mathbb{Z})}.
\end{aligned} \tag{5.7}$$

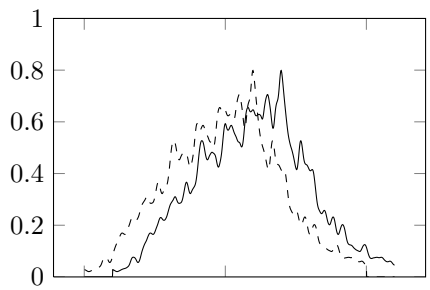
The bound (5.7) shows that the difference between a signal and a translated version of that signal grows at most linearly in the translation. Since this estimate also holds for the original signal (i.e., non-filtered version), we do not have a better complexity of the estimate for the filtered versions. But as in (5.4) we have that $\|\boldsymbol{x}_k\|_{\ell^2(\mathbb{Z})} \leq \|\boldsymbol{x}\|_{\ell^2(\mathbb{Z})}$, so even though we add several norms in the metric (5.3), the translation dependence is less severe than that of the ℓ^2-norm.

To illustrate this, consider Figure 5.2 where the maps $h \mapsto d(\boldsymbol{x}, \tau_h \boldsymbol{x})$ and $h \mapsto \|\boldsymbol{x} - \tau_h \boldsymbol{x}\|_2$ are presented.

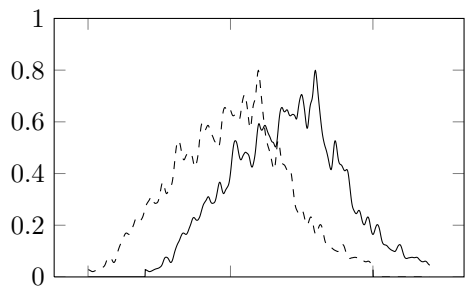
In conclusion the ℓ^2-norm suffers much worse from translation of signals than the metric (5.3). To further circumvent the translation issue in future work, one idea could be to draw inspiration from the before mentioned metric in [**27**], where the idea is to separately measure the difference in magnitude and displacement.



(a) The maps $h \mapsto d(\boldsymbol{x}, \tau_h \boldsymbol{x})$ and $h \mapsto \|\boldsymbol{x} - \tau_h \boldsymbol{x}\|_2$ for both the measured and computed signal, where d is the pseudo-metric from (5.3). In both cases the pseudo-metric outperforms the ℓ^2-norm.



(b) Translation with 100 units.



(c) Translation with 200 units.

Figure 5.2: Effect of translation on the ℓ^2-norm and the metric (5.3) along with examples of how the measured signal looks compared to versions translated 100 and 200 units.

5.2 National Gallery of Denmark

As mentioned in the introduction, I started collaborating with the National Gallery of Denmark (NGD) in the summer of 2011. All digital reproductions of paintings and drawings presented in this section are recorded at NGD. I would like to acknowledge the willingness to cooperate from the staff at NGD – data acquisition and questions from me have been resolved very quickly. This goes especially for Niels Borring, Troels Filtenborg and Jakob Skou-Hansen.

The NGD has several drawings *and* paintings attributed to both Pieter Bruegel the Elder and Jan Bruegel the Elder and I initially contacted them because I thought it would be interesting to see if I could reach conclusions about their Bruegel works.

The newspaper article [**55**] that sparked my interest in collaborating with NGD was about their success with determining that a *painting* formerly attributed to Bruegel was a forgery. The painting in question is presented in Figure 5.3a.

Paintings and drawings have a very different nature – and even though our authentication method is able to distinguish between works of different artists when looking at the two kinds separately, we do not achieve reasonable results when mixing the different mediums.

A suggestion for a workaround was to use infra red recordings of the under drawings of the paintings; using under drawings from X-rays has been used successfully by Stork and Kossolapov [**57**] to assess working methods of the Renaissance painter Lorenzo Lotto.

The under drawings from NGD are seen in Figure 5.3b and as an artistic layman I would still deem this to be very different from the drawings I have available as training material, an example of which is presented in Figure 3.1. Indeed, comparing the painting and under drawing in Figure 5.3a with our training material did not yield any reasonable results.

(a) Painting. (b) Under drawing.

Figure 5.3: The painting that NGD deemed a forgery as reported in [**55**] as well as infra red recordings of the under drawing. The copyright of the digital reproductions belongs to SMK-foto and the drawings belong to the National Gallery of Denmark. The image has NGD catalog number KMS3924.

Luckily the NGD also have drawings attributed to Bruegel, which I now have digital reproductions of; an exposition of these drawings is shown in Figure 5.4.

Applying our authentication on the digital reproductions in Figure 5.4 was not immediately successful – the problem being that the data was too different. The differences were manifested in the different recording equipment: The digital reproductions of the Bruegel drawings used for training the model are much older than the recent images from NGD. This has two significant implications: The resolution is lower and the camera has recorded much less meta data – most important is the lack of focus distance as an approximation of the distance between camera and drawing. On the other hand, the images from NGD contains such information.

As explained in Chapter 2 it is of great importance to our authentication method that the zoom level in the images are comparable; the number of pixels per area unit on the canvas should be roughly the same.

It makes good sense that this is necessary for our method since it is based on a multiresolution analysis – it does not make sense to compare the same levels of multiresolution representations if the brushstrokes in one are the size of the objects they constitute in the other.

In Chapter 2 the task of determining how much we should zoom in in the high resolution images was easy, since the two sets of images were of the same paintings. In this case we can compare distinguished features in the different digital reproductions of the same image and select the downsampling level such the features are of the same size. Even when this have to be done manually it is a simple task that can be performed by non-experts.

The situation is, of course, more difficult when we have to compare different paintings and judge how much to downsample the ones of higher resolution to match the level of the others.

The thesis that lay ground for the automatic authentication is that the brushstrokes should be consistent in a painters work. When attempting to estimate the appropriate zoom level I therefore think the brushstrokes should be compared – especially the width of brushstrokes. Even in images of high resolution typical brushstrokes are not many pixels wide and simply zooming in until a brushstrokes is dominating will blur the image too much for any recognition task by the human eye.

To circumvent this problem I have inferred the zooming level from an edge map of the digital reproduction using the method described in [**65**]. The approach does involve some manual interaction and can undoubtedly be improved, but it has been exploited as time has permitted.

In more detail, I have first used a Canny edge detector [**11**] to extract the edges at a suitable level and part of an image – as exemplified in Figure 5.5. The criterion for a being suitable in this context is that there are non-overlapping brushstrokes.

Because the width of the brushstrokes do span *some* pixels, we get a response from *both* sides of the brushstrokes with the edge detector. So the width of the brushstrokes can be estimated from the distribution of widths along such a brushstrokes. In [**65**] is presented a method for estimating *thick* lines by means of the Radon

(a) Catalog number KKS7294. Jan Bruegel.

(b) Catalog number KKSgb9871. Artist unknown.

(c) Catalog number KKSgb7860. Pieter Bruegel.

(d) Catalog number KKSgb9869. Artist unknown.

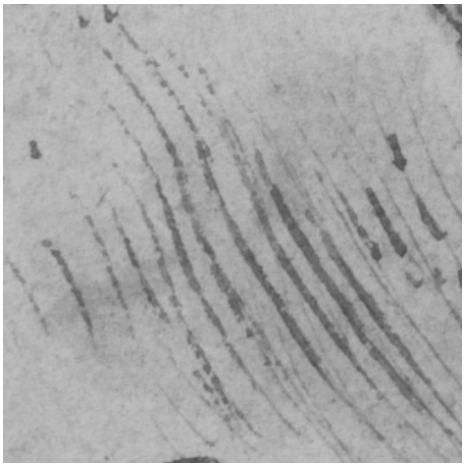
Figure 5.4: Drawings from NGD related to Bruegel. The copyright of the digital reproductions belongs to SMK-foto and the drawings belong to the National Gallery of Denmark. The NGD catalog number and artist affiliation is listed below each drawing.

(e) Catalog number KKSgb6569. Artist unknown.

(f) Catalog number KKS6568. Artist unknown.

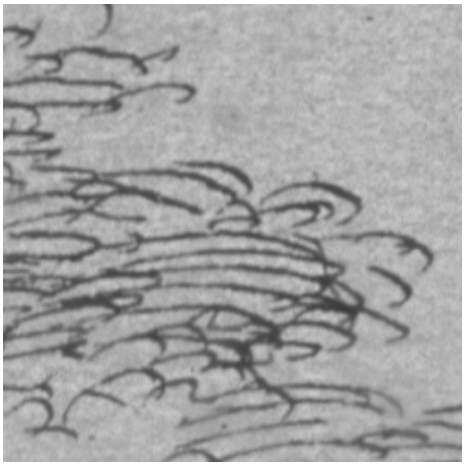
(g) Catalog number KKS7265. Jan Bruegel.

Figure 5.4: Drawings from NGD related to Bruegel. The copyright of the digital reproductions belongs to SMK-foto and the drawings belong to the National Gallery of Denmark. The NGD catalog number and artist affiliation is listed below each drawing.



(a) Patch from Figure 5.4a.

(b) Edges in (a).



(c) Patch from Figure 3.1.

(d) Edges in (c).

Figure 5.5: Patches from different Bruegel drawings along with the edges found in these patches. The two patches have been scaled differently to fit on the page, but Figure 5.4a is larger.

transform. The Radon transform along a line $L \in \mathbb{R}^2$ of a (suitable) function f is defined as the line integral

$$(Rf)(L) = \int_L f. \tag{5.8}$$

By parametrising the line with polar coordinates, the Radon transform of a function defined on $\mathbb{R}^2$ depends on the angle with the abscissa and the distance from the origin. The idea in [**65**] for utilizing the Radon transform to estimate line widths is that for a (reasonably straight) line the Radon transform will have large responses for a narrow band of distances as well as angles. The reason for assuming that the lines are thick (i.e., more than just a single pixel), is that the line width is estimated by the bandwidth of the Radon transform restricted to the dominant angle in order to reduce the noise impact. An image containing a single, filled brushstroke fits into this framework.

There is (at least) one thing to keep in mind with this approach: From a binary image with edges it is relatively straightforward to compute the connected components, which should resemble the brushstrokes; indeed, such a procedure is implemented in Matlab.

However, even though the finest edges/noise can be filtered out by choosing an appropriate threshold in the edge detection, I have still handpicked the connected components that are sufficiently large and has two sides, i.e., the parts that looks like brushstrokes. Just picking any part with brushstrokes does not guarantee good width estimates – it is problematic if too many strokes overlap.

Another issue is the filling process of the brushstroke from the edges. If the edges computed from a brushstroke form a closed curve the filling process is straightforward, but in many of the brushstrokes I have considered, this is not the case – sometimes only a few pixels are missing from a closed curve. A possible solution to this problem could be to join the edges of a brushstroke into a closed curve by e.g. the method proposed in [**52**].

In Figure 5.6 a single brushstroke from Figure 5.5 is presented along with a filled version of this brushstroke. In Figure 5.7 is the Radon transform of the filled brushstroke; it is seen that the peak of the Radon transform is well localized with respect to both angle and distance.

Estimating the width of several brushstrokes in different paintings in this fashion, it appears that the brushstrokes in the images from NGD are about twice as wide as those in our training images.

Besides choosing the right scale in the images, we know from Chapter 3 that our authentication method can benefit from increasing the contrast in the images in the right way. Especially if we make the images slightly lighter, the authentication improves.

Having tried different different contrast increasements it is concluded that for the drawings from NGD the best results are obtained when we do not alter the contrast.

At the end of day we have downsampled the images from NGD to half their original size and classified them with our training images. These preprocessing steps made the test and training images so alike that a comparison seems reasonable – as

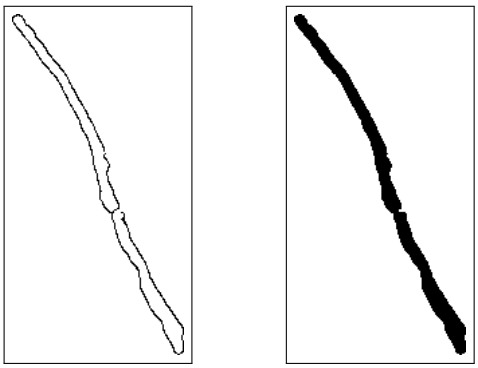


Figure 5.6: A single brushstroke from the edge map in Figure 5.5b and a filled version of this brushstroke.

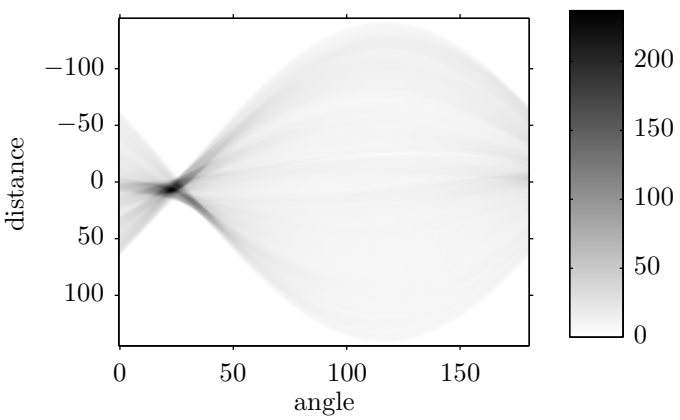


Figure 5.7: Radon transform of the filled brushstroke from Figure 5.6.

illustrated by the multidimensional scaling embedding of the pairwise distances between the images in Figure 5.8.

From Figure 5.8 we also see that the drawings from NGD are arranged in satisfactory manner: The drawing attributed to Pieter Bruegel is closer to the training Bruegels, the drawings attributed to Jan Bruegel are further away from the training Bruegels and the images whose attribution is unknown lies in between.

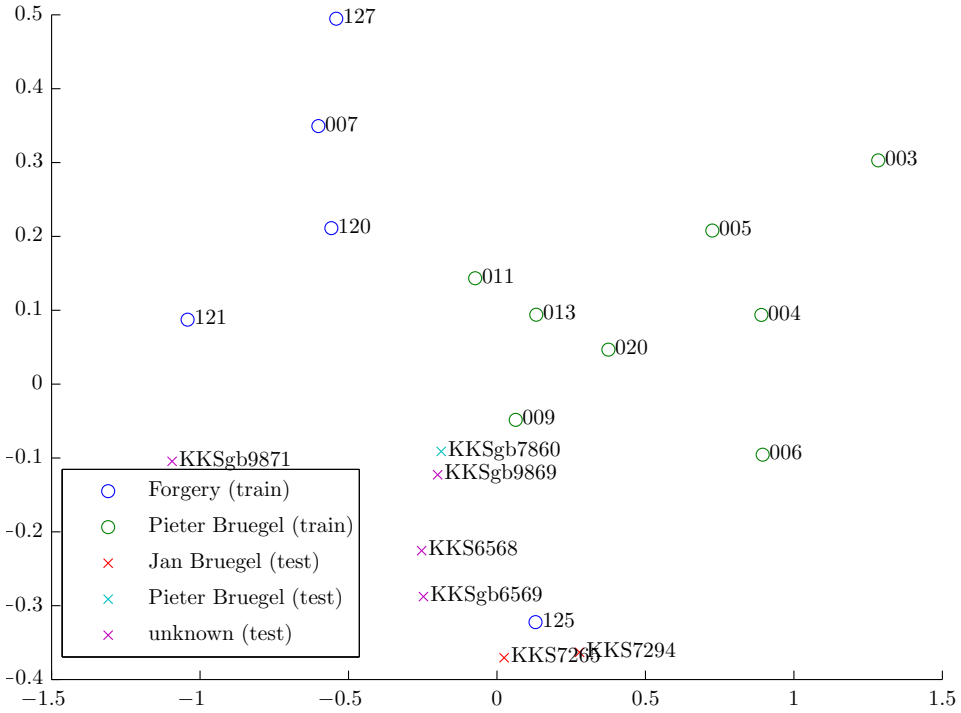


Figure 5.8: Illustration of the classification of Bruegel drawings from NGD by an embedding of points respecting the pairwise distances between the models of the different images.

Further details about the classification can be found in Table 5.1. Here it is seen that with a three nearest neighbour classifier, all but one of the images from NGD are classified as Pieter Bruegel drawings. However, it is only the authentic drawing by Pieter Bruegel and one unclassified drawing where all three of its nearest neighbours are authentic, for the rest of the drawings one or more of the forgeries are also close.

5.2.1 Sparse coding

The authentication method in [**35**] based on sparse coding was tested on the very same set of Bruegel images used as a training set here, so it seems natural to employ

drawing	neighbours	category	distance	artist
KKS7294	125	F	0.1369	Jan Bruegel
	020	A	0.3997	
	009	A	0.4331	
KKSgb7860	011	A	0.2196	Pieter Bruegel
	009	A	0.2867	
	013	A	0.3406	
KKS7265	125	F	0.1403	Jan Bruegel
	009	A	0.2924	
	013	A	0.3368	
KKSgb9869	011	A	0.2593	Unknown
	009	A	0.3044	
	013	A	0.3523	
KKSgb6569	011	A	0.3328	Unknown
	125	F	0.3676	
	009	A	0.3957	
KKS6568	011	A	0.2770	Unknown
	125	F	0.3462	
	009	A	0.3685	
KKSgb9871	121	F	0.2810	Unknown
	007	F	0.5570	
	120	F	0.5845	

Table 5.1: The Bruegel drawings from NGD used in the experiments in this section. For each drawing is listed the name of the three nearest neighbours amongst the training images, the category of each of these training images, the distance to each of the nearest neighbours and finally the artist of the drawing.

the sparse coding method on the drawings from NGD.

The method from [**35**] was sketched in Section 1.1, but I will elaborate it here. The sparse coding framework developed in [**49**, **50**] aims at learning an *adaptive* frame of functions from images of natural scenes. In [**28**] it was demonstrated that visual art shares many of the statistics found in natural scenes and it therefore makes sense to learn sparse coding bases from digital reproductions of drawings and paintings. When the frame is learned from an image, it must in some sense provide an optimal representation of this image – and the criterion used for optimality is sparsity. This means that a typical patch from an image can be represented with very few functions from the learned sparse coding frame; the specific functions needed for a representation may vary with the different patches, but only few functions are needed for a given representation.

In [**35**] sparse coding was utilized for authentication in the following manner: First a sparse coding basis is learned from the collected set of authentic training images – this ensures that every authentic training image can be represented sparsely in this basis. To determine if a test image is authentic we pick random patches from this image as well as patches from known forgeries and represent all patches in the learned sparse coding basis. If the patches from the test image are represented more sparsely in the authentic sparse coding basis than the patches from the forgery, we conclude that the test image is likely authentic. If, on the other hand, the test image is not represented more sparsely than the known forgery, we cannot conclude that it is authentic.

By performing this experiment a number of times we can supply our conclusion with confidence estimates: If the test image is not more like the known authentic images than the known forgeries, then, on average, the test image will be represented at least as bad as the forgeries in the sparse coding basis from the authentic images. Under this hypothesis the number of times the test image is represented more sparsely in the authentic sparse coding basis is binomially distributed with probability parameter $1/2$. By repeating the test explained above, we can assign confidence intervals to the decision about the category of the test image.

There are many popular ways of measuring sparsity; in [**35**] sparsity of the representation of a patch is estimated by the kurtosis of the coefficients used to represent the patch in the sparse coding basis. Kurtosis seems to be a reasonable proxy for measuring sparsity in this context, as all but one of the authentic drawings are classified as authentic in a leave-one-out cross-validation. However, as discussed in [**48**], kurtosis might not be appropriate in all situations.

Applying the sparse coding method on the drawings from NGD it is concluded that *all* drawings are more likely authentic than the known forgeries – and the conclusion is highly significant. This conclusion is not comforting, but it must be kept in mind that it might be true: It is possible that the drawings from NGD are indeed more like the authentic Pieter Bruegel drawings than the known forgeries we test them against.

CHAPTER 6

Epilogue

6.1 Discussion

In this section I will discuss some of the pros and cons of different methods for artist authentication – including the one presented in Chapter 2. In Section 2.5 we did discuss our method, but more can be said.

Two authentication methods have been utilized in the present thesis; the method developed as part of this PhD project and the method presented in [**35**] that uses sparse coding.

In my opinion the two methods approach the authentication problem from opposite sides:

- The method from Chapter 2 aims at learning a decision rule that can separate known forgeries from known authentic images and obtains a description of features that are distinctive for the authentic images. Therefore, if it is realistic to correctly classify authentic images, since they belong to a well defined and well described class.
- The method from [**35**] bases its conclusion on the misfit of the known forgeries. If a test image fit the authentic characteristics significantly better than the known forgeries, it is also considered authentic – without further comparison to the known authentic images. A benefit of this approach is that it is likely to identify forgeries that are less like the known forgeries in the ground truth available.

It is important to be able to identify both classes; authentic paintings should be identified as such and likewise with forgeries. In this regard it is of interest to apply methods suited for both tasks.

However, there are practical considerations that must be taken into account and remembered when interpreting the results.

The contourlet-based method from Chapter 2 depends on ranking the parameters from the hidden Markov trees fitted to the contourlet transforms. This ranking is performed with a suitable regression and as mentioned the high number of parameters and relatively small data sets makes makes a regular regression infeasible. Instead, we employ regression with regularization in form of lasso regression. The lasso has the downside that we do not get confidence bounds on the parameters estimated, leaving us to trust the point estimates alone.

It is unfortunate that the sparse coding method from [**35**] is not successful in classifying the drawings from NGD. However, the sparse coding method has several intrinsic assumptions about the data that can influence the conclusions it offers and such dependencies need to be investigated further. Amongst the dependencies can be mentioned the procedure for choosing patches, preprocessing of the images and the measure of sparsity used.

6.2 Conclusion

In the preceding chapters the main part of this thesis has been presented: An approach for authentication of paintings and drawings from digital reproductions and statistical modelling of artistic style.

Using mathematical methods to assist the task of authenticating paintings and drawings is fairly new, but several different approaches have been suggested by different authors. In Chapter 2 I have introduced the main method developed for authentication during my PhD studies. The authentication method has some parts in common with other authentication approaches, but a new aspect presented in the current work is a study of how well our authentication methods performs in the presence of disturbing factors. The latter has proven to be very important in practice; an automatic authentication method is of little interest if it is more cumbersome than doing the work "by hand". As demonstrated in Chapter 3, our authentication method is robust under a variety of factors that are likely to interfere with the digital reproductions.

Digital reproductions of visual art can be used for more than authentication as demonstrated in Chapter 4 where we investigated the use of higher order statistics for a broader range of categorizations.

6.3 Perspectives

From the experiments performed with digital artist authentication the future for this topic looks promising. It is not likely that automatic methods will replace art experts anytime soon, but they can aid decisions in a presorting with consistent results.

I think that there are many paths to be exploited to make the automatic au-

thentication methods a more reliable – and trusted – tool. Let me mention some important aspects that I think it would be relevant to investigate.

6.3.1 Cracks

A significant imprint from time upon oil paintings is the cracks that occur in the paint. When the cracks becomes sufficiently dominant it is realistic that they interfere with an automatic authentication procedure. For the method we have developed based on a multiresolution decomposition, large cracks will stand out as significant high frequency content, which is also the content we use for authentication.

I know of attempts to remove cracks in the digital reproductions of paintings [**18**] by means of *inpainting*. Inpainting is an interesting topic that aims at filling out empty or unwanted parts of an image by using the information available in the remainder of the image. There are different approaches to inpainting and inpainting have proven to be usefull in visual stylometry [**38**].

Inpainting methods are at present capable of producing visibly plausible results, where the human eye cannot distinguish between original image content and the parts that have been inpainted. However, as demonstrated in [**62**], some inpainting methods *do* introduce patterns that can be detected by other means. As far as I know, it has not been investigated if inpainting affects authentication methods.

The inpainting is only one part of the work needed for filling out cracks; first the cracks must be detected in a reliable manner. In [**18**] the cracks are detected using the top hat transform. A potential method for crack detection is to apply the beamlet transform [**22**]; the beamlet transform is a multiresolution transform that can be used to find lines at different orientations and scales in images – and it has proven to work at finding cracks in pavement [**63**].

As a side note it should be mentioned that cracks are not only disturbing – they can also be used for classification as demonstrated by Spike Bucklow in e.g. [**7**].

6.3.2 Canvas

When the artist have painted on a coarse canvas and the layers of paint are not too thick, the canvas can severely influence the algorithms searching for structure.

An example of this are the paintings by Charlotte Caspers [**12**] included in the experiments in Chapter 2 and 4, that are painted on a variety of canvases – including very coarse bare linen canvas as exemplified in Figure 6.1.

The paintings from [**12**] are made in pairs – an original and a copy made by the same artist. As reported in Chapter 2 we could not separate the originals from the copies in this data set – in fact the images clustered in pairs by the materials and paint used for producing them. Different combinations of canvas, paint and brushes were used and a reasonable explanation is that the canvas is too dominant – like for paintings in Figure 6.1. This hypothesis is further supported by the experiments performed in connection with the work from Chapter 4: When training a sparse

Figure 6.1: Part of painting by Charlotte Caspers from [**12**] on bare linen canvas.

coding basis from images such as Figure 6.1 the canvas had a significant imprint on a majority of the basis function that resembled the canvas structure.

Thus, if we wish to draw conclusions from the actual painted parts, it it necessary to compensate for the influence of the canvas. The canvas (usually) have a periodic structure, which suggests that the impact can be analyzed and maybe removed using Fourier analysis. However, it must be taken into account that the influence of the canvas is less severe in the parts with thick layers of paint.

6.3.3 Scaling

In both Chapter 2 and Section 5.2 it was concluded that in order to make a sensible comparison between different paintings or drawings they need to be of comparable sizes in their digital reproductions.

Determining the right zoom level is not a trivial task and it is desirable to develop methods to (further) automate the process of finding the appropriate zoom level.

An alternative would be to base (part of) the authentication task on scale-invariant features. For both tasks an involvement of methods based on scale-space (see e.g. [**43**]) might be a lead-in.

Such an approach has recently been successful in distinguishing between distinct writing style in ancient manuscripts [**26**]. An approach for authentication of paintings in this regard might be to extract individual brushstrokes as suggested in Section 5.2 and characterize the distribution of brushstrokes in a painting.

6.3.4 Standard for data collection

Above all, the lack of a standard for collecting the digital reproductions the authentications are based on constitutes a problem for reproducibility. Data material for experiments is scarce and not easy to come by – often with obligations not to share or redistribute; so simply abiding by the guidelines of reproducible research

and sharing code and data is not always an option.

Furthermore, if an authentication method is only successful in separating a single artist from attempted forgeries, this has to be a very important artist to make the method of general interest. So when testing an authentication approach on a data set that is different from the one for which the method is developed, this data should comparable. As demonstrated in this thesis, realistic comparability is not possible if the different data sets have been obtained in significantly different ways – thus the need for a collection standard arises.

The definition of a standard is at current more wishful thinking – at least until there is a better understanding of the factors influencing the decisions from an automatic authentication method.

APPENDIX A

Contourlets

The contourlet transform has been used extensively throughout this thesis and was briefly introduced in Section 2.3.1.

Here I will give a more thorough treatment of the contourlet transform; the emphasis will be on the discrete contourlet transform and not on the continuous counterpart. The reason for this choice is that I wish to provide insight into our motivation for using the contourlet transform on the digital images we have used for experiments – the joyful activities of *proving* that the contourlet transform behaves well has not been my purpose and thus I do not think that involving the continuous contourlet functions will contribute with further understanding. If the reader seeks enlightenment in this regard, the original article [**21**] can be recommended.

Initially I will try to motivate the need and behavior of a good transform for digital images.

When representing images (and signals in general) in a basis, we wish to do so *efficiently*. Furthermore, as has been the focus of my work, when analysing images it is desirable to have an image representation that emphasize the patterns that you are looking for.

Effectiveness of a representation is measured by how *sparse* the representation is, that is, how few functions from the basis are necessary for the representation of a "nice" object? To quantize this we study the behavior of the error rate from *M-term approximations* of the images: Let $\{\phi_i\}_{i\in I}$ be a frame for an appropriate Hilbert space where our image f lives, such that we can decompose the image as

$$f = \sum_{i\in I} c_i \phi_i. \tag{A.1}$$

The expansion in a frame is not necessarily unique, but one way to ensure that the

expansion is obtained consistently is to use a canonical frame expansion – this is also the approach with contourlets. Given an expansion as in (A.1), the associated M-term approximation of f is defined to be

$$\widehat{f}_M = \sum_{i \in I_M} c_i \phi_i, \tag{A.2}$$

where I_M is the set of indices of the M numerically largest coefficients from (A.1).

A very distinctive feature of images is the *edges* – and in most images edges occur at a variety of scales and orientations. Edges are the high frequency content in an image and thus the parts that are difficult to represent; in parts with low frequency content the image is more homogeneous and therefore easier to represent. Thus to find an efficient representation for images, it is necessary to find efficient representations of the edges.

The human visual system is well suited for the task of processing edges, since the receptive fields in the visual cortex are localized, orientated and bandpass [**49**]. The search for efficient transforms for images have therefore attempted to match this behavior with the basis functions, which is achieved by the contourlet frame functions.

In [**23**] it was shown that if the (analog) image is a C^2-function, except possibly along piecewise C^2-curves, then the best possible M-term approximation decays at a rate proportional to M^{-2}. The contourlet transform almost achieves this rate, since the M-term approximation associated with the canonical frame expansion from the contourlet frame obeys

$$\|f - \widehat{f}_M\|_2^2 \lesssim (\log M)^3 M^{-2}. \tag{A.3}$$

Transforms such as curvelets [**10**] and shearlets [**39**] achieves the same rate, but to my knowledge no transform has been found that outperforms the rate in (A.3).

The curvelet transform was devised prior to the contourlet transform, so one may be inclined to ask if other transforms are necessary? The answer to this question depends on the applications intended for the transform. If we work in the continuous case, i.e., on a function space like $L^2(\mathbb{R}^2)$, then the curvelet transform is well suited for the task. If, on the other hand, we work with discrete data, the curvelet transform has some limitations, since it is adapted to the discrete case via a sampling procedure in the frequency domain, see e.g. [**9**].

Do and Vetterli's motivation for creating the contourlet transform was the need for a transform that is *designed* to work on discrete data, with the benefits and limitations this structure imposes. In applications the construction of the contourlet transform also benefits from the fact that it is implemented through a double filter bank (something that is not possible for the curvelet transform) and this ensures a fast implementation

As a side note I want to mention that the newer shearlet transform works equally well on continuous and discrete data, similar to the behavior of the wavelet transform. However, my acquaintance with the shearlet transform was initiated when my work with contourlets for authentication as well as this project was almost complete, so I have not investigated this thoroughly.

A.1 Contourlet transform

Let us move on to the construction of the contourlet transform.

The contourlet transform is a combination of two transforms, namely

1) The Laplacian Pyramid [**8**] with special filters and reconstruction scheme [**20**] for dividing the image into a low pass and a high pass subband.

2) A two dimensional directional filter bank (originally introduced in [**2**] and further developed in [**19**, **21**]) for splitting the high pass subband into different spatial frequency orientations.

The combined properties of these transforms provide us with the desired properties of a multi directional transform for digital images – and by successively applying the transform on the low pass image, we also achieve a multiresolution representation.

To explain how the low pass image is computed, we need the notion of sampling in two dimensions. A sampling lattice associated with an invertible matrix $M : \mathbb{Z}^2 \to \mathbb{Z}^2$ is the subset of $\mathbb{Z}^2$

$$\mathrm{LAT}(M) = \{\boldsymbol{n} \,|\, \boldsymbol{n} = M\boldsymbol{m}, \boldsymbol{m} \in \mathbb{Z}^2\}.$$

and downsampling by a matrix M is defined as

$$((\downarrow M)\boldsymbol{x})[n] = x[M\boldsymbol{n}], \quad n \in \mathbb{Z}^2.$$

Upsampling by a matrix M from a lattice LAT is defined by

$$((\uparrow M)\boldsymbol{x}) = \begin{cases} x[M^{-1}\boldsymbol{n}], & M^{-1}\boldsymbol{n} \in \mathrm{LAT} \\ 0, & \text{otherwise.} \end{cases}$$

The diagram in Figure A.1 illustrates the double filter bank of the contourlet transform.

Laplacian Pyramid In the first part of the contourlet transform we apply a low pass filter to our image, yielding a coarse approximation of the image. By downsampling the coarse approximation we obtain the low pass subband and by subtracting the coarse approximation from the original image, we obtain the high pass subband – as illustrated with the filter bank in Figure A.2a.

The Laplacian Pyramid has two distinctive features that are useful in the current context.

First of all there is only one high pass subband of the image – as opposed to e.g. the wavelet transform, where the number of high pass subbands increases with the dimension of the signal.

Secondly, when we do not downsample the high pass subband, we ensure that it is indeed a high pass subband. To explain this, consider for simplicity the situation

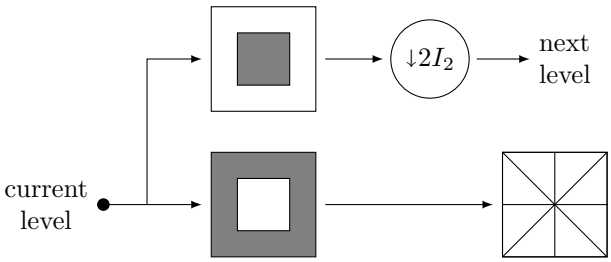


Figure A.1: Filter bank of the contourlet transform: At each resolution level the image is decomposed into a low frequency subband which is downsampled for use in the next level and a high frequency subband which is partitioned into directional subbands. Downsampling is performed with two times the two dimensional identity matrix I_2, i.e., downsampling by a factor two in both horizontal and vertical direction.

in one dimension. A combination of the Fourier operator $\mathcal{F} : \ell^2(\mathbb{Z}) \to L^2([-\pi, \pi])$ and a downsampling operator $(\downarrow 2) : \ell^2(\mathbb{Z}) \to \ell^2(\mathbb{Z})$ on a signal $\boldsymbol{x} \in \ell^2(\mathbb{Z})$ yields

$$\begin{aligned}\big(\mathcal{F}(\downarrow 2)\boldsymbol{x}\big)(2\xi) &= \sum_{k\in\mathbb{Z}} x[2k]e^{-i2k\xi} = \sum_{n\in\mathbb{Z}} x[n]\frac{1+(-1)^n}{2}e^{-in\xi} \\ &= \frac{1}{2}\sum_{n\in\mathbb{Z}} x[n]\big(e^{-in\xi} + (-e^{-i\xi})^n\big) \\ &= \frac{1}{2}\sum_{n\in\mathbb{Z}} x[n]\big(e^{-in\xi} + e^{-in(\xi+\pi)}\big) \\ &= \frac{1}{2}\big(\widehat{\boldsymbol{x}}(\xi) + \widehat{\boldsymbol{x}}(\xi+\pi)\big). \end{aligned} \tag{A.4}$$

So the downsampling operator stretches the spectrum of a signal and shifts it. Shifted (and possibly stretched) versions of a signal are called *aliased* versions [**60**].

Let m_1 be a high pass filter and $\boldsymbol{g}$ be the corresponding coefficient sequence; that is, $\widehat{\boldsymbol{g}}(\xi) = \sqrt{2}m_1(\xi)$. Using (A.4) on the output from the downsampled high pass filter we get that

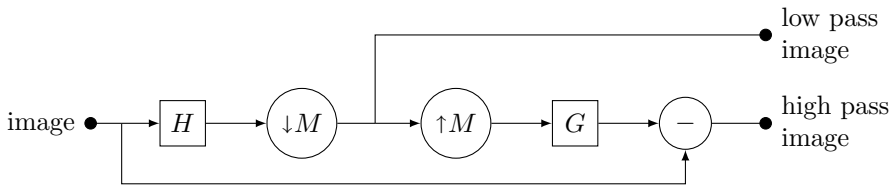
$$\begin{aligned}\big(\mathcal{F}(\downarrow 2)(\boldsymbol{g} * \boldsymbol{x})\big)(2\xi) &= \frac{1}{2}\big(\widehat{\boldsymbol{g}}(\xi)\widehat{\boldsymbol{x}}(\xi) + \widehat{\boldsymbol{g}}(\xi+\pi)\widehat{\boldsymbol{x}}(\xi+\pi)\big) \\ &= \frac{1}{\sqrt{2}}\big(m_1(\xi)\widehat{\boldsymbol{x}}(\xi) + m_1(\xi+\pi)\widehat{\boldsymbol{x}}(\xi+\pi)\big). \end{aligned} \tag{A.5}$$

The last term in (A.5), $m_1(\xi+\pi)\widehat{\boldsymbol{x}}(\xi+\pi)$, is an aliased version of the first term, which is a regular high pass filtered version of $\boldsymbol{x}$. When shifting the high frequency content, it is going to intersect with the low frequency content – which was the very content we filtered out.

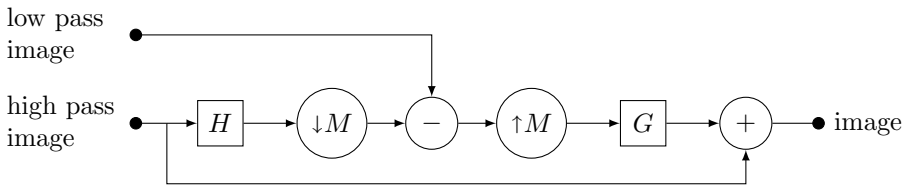
A small downside of the Laplacian Pyramid is that since we do not downsample the high pass subband the representation is redundant.

The original way of reconstructing an image from the two subbands is by upsampling the low pass subband, reverse the filtering and add the high pass subband. This method is straightforward and at first sight it seems to be a suitable inverse for the decomposition.

However, it was shown in [**20**] that the original reconstruction scheme is sub-optimal and in the same paper a better way of reconstructing the image was proposed using orthogonal filters. This new reconstruction is utilized in the contourlet transform and works as shown in Figure A.2b: The low pass synthesis filter is applied to the high pass subband and the result is downsampled and subtracted from the low pass subband. This mixed image is upsampled and the high pass synthesis filter is applied to it. Finally the high pass subband is added to yield the reconstruction.



(a) Decomposition.



(b) Reconstruction.

Figure A.2: Diagram of the Laplacian Pyramid decomposition and reconstruction as performed in the contourlet transform. H is the low pass filter and G is the high pass filter.

Directional filter bank Once we have obtained the multiscale decomposition of an image with the Laplacian Pyramid, we utilize a directional filter bank on each of the high pass subbands.

The directional filter bank is based on the work in [**2**], where the authors construct a filter bank that partitions the spatial frequency domain into wedge shaped parts of equal phase, as illustrated in Figure A.3.

This directional filter bank can be maximally decimated and achieve perfect reconstruction.

In [**19**, **21**] it was shown how to construct a two-channel filter bank where the response in the spatial frequency domain of the channels are the horizontal and

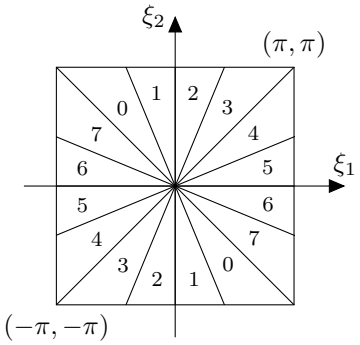
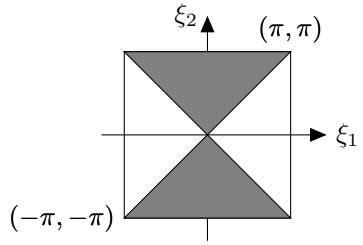


Figure A.3: Possible partition of the spatial frequency domain by the directional filter bank along with the ordering of directions used in the contourlet filter bank.

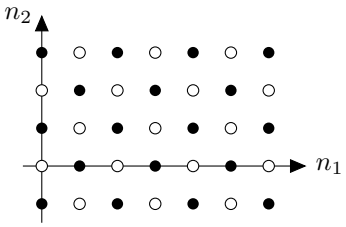
vertical directions. This filter bank is obtained by applying an hourglass filter from [**2**] with spatial frequency response as in Figure A.4a in combination with a quincunx sampling lattice (see e.g. [**60**] for a thorough introduction to the appealing features of this sampling lattice).

The word quincunx stems from the latin words for five (quinque) and one (unus) and is used to describe the five pattern from a die ⚄ [**17**] and more generally an arrangement of five objects with four at the corners of rectangle and the fifth at its center. A quincunx lattice is a lattice that resembles the geometric structure of a quincunx, as illustrated in Figure A.4b and such a lattice can for instance be obtained from the sampling matrix

$$Q = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}. \tag{A.6}$$



(a) Idealized support of the hourglass filter from [**2**].



(b) Example of a quincunx lattice.

Figure A.4: Ingredients in one half of the directional filter bank: A filter whose support in the spatial frequency domain prefers horizontal directions and quincunx sampling lattice. Gray indicates the supported parts.

Applying such filter banks gives a coarse partition of spatial frequency directions in mainly vertical and mainly horizontal high frequency content. To obtain partitions

at a finer directional granularity we combine the hourglass filters with *shearing* operators.

An image is sheared horizontally by a matrix of the form

$$\begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}$$

with $k > 1$; for vertical shears we transpose this matrix. By choosing k to be an integer, the shearing matrices are sampling matrices and they are furthermore bijective from $\mathbb{Z}^2$ to their associated sampling lattice.

After shearing an image, straight lines become skew and vice versa, so applying the quincunx filter bank on a sheared image will return parts that used to belong to a tilted directional subband. By using an inverse shearing after filtering, the acquired "straight" lines will be brought back to their original direction.

Contourlet filter bank The combination of a Laplacian and a directional filter bank results in the contourlet filter bank. Here is an explanation of what the frequency response of the contourlet transform at a particular level and direction looks like.

Consider the high pass output $\boldsymbol{y}$ from level N of the contourlet filter bank, where N is not the finest high pass level, nor the low pass level. Let $\boldsymbol{h}$ denote the low pass filter, $\boldsymbol{g}$ denote the high pass filter and $\boldsymbol{d}$ be the directional filter. Convolution with $\boldsymbol{h}$ in a filter bank is denoted $\widehat{\boldsymbol{h}}(\boldsymbol{\xi})$ and likewise for $\boldsymbol{g}$. The image $\boldsymbol{x}$ is transformed into $\boldsymbol{y}$ through the filter bank

$$\boldsymbol{x} \to \boxed{\widehat{\boldsymbol{h}}(\boldsymbol{\xi}) \to \textcircled{\downarrow 2}}_N \to \widehat{\boldsymbol{g}}(\boldsymbol{\xi}) \to \widehat{\boldsymbol{d}}(\boldsymbol{\xi}) \to \boldsymbol{y}, \tag{A.7}$$

where the box is applied N times. To rewrite this filter bank we need a multirate identity [**61**] valid for general filters $\boldsymbol{g}$:

$$\boldsymbol{x} \to \textcircled{\downarrow k} \to \boldsymbol{g}(\boldsymbol{\xi}) \to \boldsymbol{y} \iff \boldsymbol{x} \to \boldsymbol{g}(k\boldsymbol{\xi}) \to \textcircled{\downarrow k} \to \boldsymbol{y}. \tag{A.8}$$

Using (A.8) we can rewrite (A.7) to

$$\boldsymbol{x} \to \widehat{\boldsymbol{g}}(2^N\xi) \prod_{i=0}^{N-1} \widehat{\boldsymbol{h}}(2^i\xi) \to \textcircled{\downarrow 2^N} \to \widehat{\boldsymbol{d}}(\boldsymbol{\xi}) \to \boldsymbol{y} \tag{A.9}$$

The first filter in (A.9) is the effect of the Laplacian pyramid. With ideal low and high pass filters as in Figure A.1, the intersection of the supports in the Laplacian filter is just that of the dilated $\widehat{\boldsymbol{g}}$. Thus we can illustrate (A.9) in *parallel from* as in Figure A.5.

Let $\boldsymbol{L}$ denote the filter from the Laplacian pyramid in (A.9). Another application of (A.8) yields

$$\boldsymbol{x} \to \boldsymbol{L} \to \widehat{\boldsymbol{d}}(2^N\boldsymbol{\xi}) \to \textcircled{\downarrow 2^N} \to \boldsymbol{y}, \tag{A.10}$$

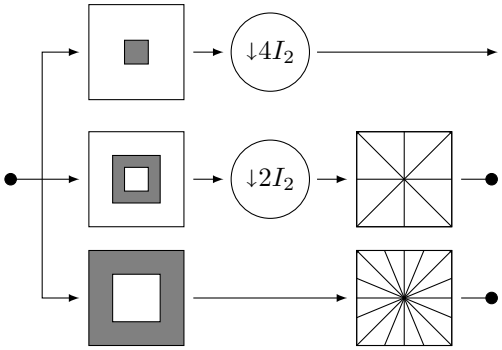
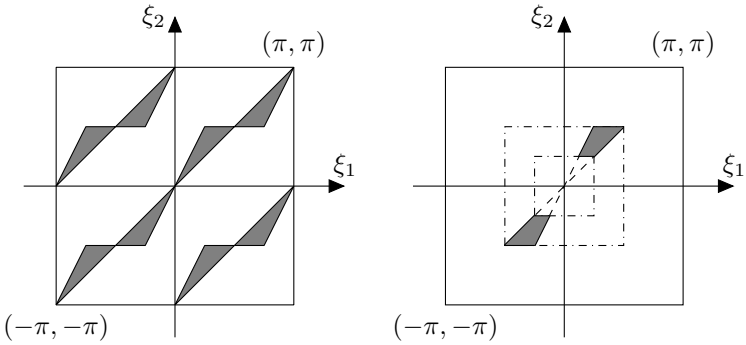


Figure A.5: Three levels of Figure A.1 in parallel form (A.9).

where the combination of the first two filter gives the final frequency response of the contourlet transform at level N and direction specified by $\boldsymbol{d}$.

Combining the parallel form Figure A.5 and the direction filter bank in Figure A.6 the frequency response of the contourlet filter bank (A.10) is visualized in Figure A.6.



(a) Upsampled directional filter.

(b) Frequency response of contourlet transform.

Figure A.6: Ideal frequency response of the representation of the contourlet transform in (A.10) at level 2 (referring to (A.7) and Figure A.5) and direction 3 (as in Figure A.3). This frequency response appears as the multiplication of the individual responses of the Laplacian pyramid at the level 2 and the upsampled directional filter bank.

Finally, when visualizing a contourlet transform, the coefficients from each directional subband as well as the low frequency approximation are displayed, as can be seen in Figure 2.2b. The ordering of the directional subbands in the frequency domain is seen in Figure A.3 and in the visualization of the corresponding transform

like Figure 2.2b, the ordering is as shown in Figure A.7.

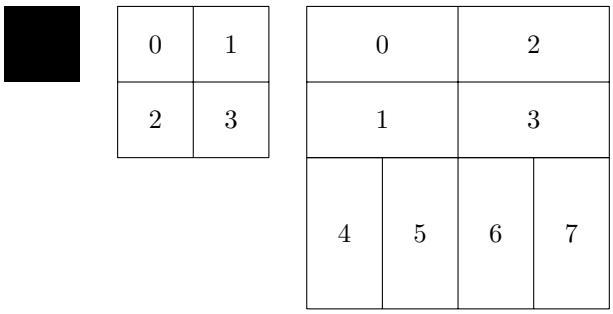


Figure A.7: With the ordering of the directional subbands as in Figure A.3 the subbands of a contourlet transform with 4 and 8 directions in the coarse and fine high pass levels, respectively. The black box represents the low pass subband. Such a transform is illustrated in Figure 2.2b.

Bibliography

[1] Patrice Abry, Herwig Wendt, and Stéphane Jaffard. "When Van Gogh meets Mandelbrot: Multifractal classification of painting's texture". In: *Signal Processing* (2012). DOI: 10.1016/j.sigpro.2012.01.016.

[2] Roberto H. Bamberger and Mark J. T. Smith. "A Filter Bank for the Directional Decomposition of Images: Theory and Design". In: *IEEE Transactions on Signal Processing* 40.4 (Apr. 1992), pp. 882–893. DOI: 10.1109/78.127960.

[3] Igor Berezhnoy, Eric Postma, and H. Jaap van den Herik. "Authentic: Computerized Brushstroke Analysis". In: *IEEE International Conference on Multimedia and Expo*. 2005. DOI: 10.1109/ICME.2005.1521739.

[4] Igor E. Berezhnoy, Eric O. Postma, and H. Jaap van den Herik. "Automatic extraction of brushstroke orientation from paintings". In: *Machine Vision and Applications* 20.1 (Jan. 2009), pp. 1–9. DOI: 10.1007/s00138-007-0098-7.

[5] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2007. ISBN: 978-0-387-31073-2.

[6] Hugh Brigstocke, ed. *The Oxford Companion to Western Art*. Oxford University Press, 2001.

[7] Spike L. Bucklow. "A Stylometric Analysis of Craquelure". In: *Computers and the Humanities* 31.6 (1998), pp. 503–521. JSTOR: 30204782.

[8] Peter J. Burt and Edward H. Adelson. "The Laplacian Pyramid as a Compact Image Code". In: *IEEE Transactions on Communications* COM-31.4 (Apr. 1983), pp. 532–540. DOI: 10.1109/TCOM.1983.1095851.

[9] Emmanuel Candès, Laurent Demanet, David Donoho, and Lexing Ying. "Fast discrete curvelet transform". In: *Multiscale Modeling & Simulation* 5.3 (2006), pp. 861–899. DOI: 10.1137/05064182X.

[10] Emmanuel J. Candès and David L. Donoho. "New Tight Frames of Curvelets and Optimal Representations of Objects with Piecewise C^2 Singularities". In: *Communications on Pure and Applied Mathematics* 57 (Feb. 2004), pp. 219–266. DOI: 10.1002/cpa.10116.

[11] John Canny. "A Computational Approach to Edge Detection". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-8.6 (Nov. 1986), pp. 679–698. DOI: 10.1109/TPAMI.1986.4767851.

[12] Charlotte Caspers. *Caspers Data Set*. URL: http://www.math.princeton.edu/ipai/datasets.html.

[13] Ian Chilvers, ed. *The Oxford Dictionary of Art*. Oxford University Press, 2004.

[14] Ian Chilvers, ed. *The Oxford Dictionary of Art and Artists*. Oxford University Press, 2009.

[15] Ian Chilvers and John Glaves-Smith, eds. *A Dictionary of Modern and Contemporary Art*. Oxford University Press Inc., 2009.

[16] Michael Clarke and Deborah Clarke, eds. *The Concise Oxford Dictionary of Art Terms*. Oxford University Press Inc.

[17] John H. Conway and Richard K. Guy. *The Book of Numbers*. Springer, 1996. ISBN: 978-0-387-97993-9.

[18] Bruno Cornelis et al. "Digital Painting Analysis, At the Cross Section of Engineering, Mathematics and Culture". In: *Proceedings of EUSIPCO*. 2011.

[19] Minh N. Do. "Directional Multiresolution Image Representations". PhD thesis. École Polytechnique Fédérale de Lausanne (EPFL), Oct. 2001. DOI: 10.5075/epfl-thesis-2500.

[20] Minh N. Do and Martin Vetterli. "Framing Pyramids". In: *IEEE Transactions on Signal Processing* 51.9 (Sept. 2003), pp. 2329–2342. DOI: 10.1109/TSP.2003.815389.

[21] Minh N. Do and Martin Vetterli. "The Contourlet Transform: An Efficient Directional Multiresolution Image Representation". In: *IEEE Transactions on Image Processing* 14.12 (Dec. 2005), pp. 2091–2106. DOI: 10.1109/TIP.2005.859376.

[22] David Donoho and Xiaoming Huo. "Beamlets and multiscale image analysis". In: *Multiscale and Multiresolution Method*. Vol. 20. Lecture Notes in Computational Science and Engineering. Springer, 2002, pp. 149–196. DOI: 10.1007/978-3-642-56205-1_3.

[23] David L. Donoho, Martin Vetterli, Ronald A. DeVore, and Ingrid Daubechies. "Data Compression and Harmonic Analysis". In: *IEEE Transactions on Information Theory* 44.6 (Oct. 1998), pp. 2435–2476. DOI: 10.1109/18.720544.

[24] Richard O. Duda, Peter E. Hart, and David G. Stork. *Pattern Classification*. 2nd ed. Wiley, 2001.

[25] *Fingerprint may lead to new da Vinci discovery*. Oct. 2009. URL: http://www.usatoday.com/news/world/2009-10-14-new-da-vinci-fingerprint_N.htm.

[26] Angelika Garz. *Efficient Layout Analysis of Ancient Manuscripts Using Local Features*. Tech. rep. Computer Vision Lab: Vienna University of Technology, Sept. 2011. URL: http://caa.tuwien.ac.at/cvl/research/tr/pdf/tr7.pdf.

[27] Thomas L. Geers. "An Objective Error Measure for the Comparison of Calculated and Measured Transient Response Histories". In: *The Shock and Vibration Bulletin* 54 (1984), pp. 99–107.

[28] Daniel J. Graham and David J. Field. "Statistical regularities of art images and natural scenes: spectra, sparseness and nonlinearities". In: *Spatial Vision* 21 (1 2007), pp. 149–164. DOI: 10.1163/156856808782713771.

[29] Daniel J. Graham, James M. Hughes, Helmut Leder, and Daniel N. Rockmore. "Statistics, vision, and the analysis of artistic style". In: *Wiley Interdisciplinary Reviews: Computational Statistics* 4.2 (Mar. 2012), pp. 115–123. DOI: 10.1002/wics.197.

[30] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. 2nd ed. Springer, 2008. ISBN: 978-0-387-84857-0.

[31] Eugino Hernández and Guido Weiss. *A First Course on Wavelets*. CRC Press, 1996.

[32] Bill Hinchberger. *Portinari Project: The Art of the Scientific Method*. Mar. 4, 2002. URL: http://www.brazilmax.com/news3.cfm/tborigem/fe_artcultmus/id/22.

[33] Warren Hoge. "Brazil Gathers Archive On Its Painter, Portinari". In: *New York Times* (May 30, 1983). URL: http://www.nytimes.com/1983/05/30/arts/brazil-gathers-archive-on-its-painter-portinari.html.

[34] Thomas Hoving. *False Impressions: The Hunt for Big-Time Art Fakes*. Simon & Schuster, 1996.

[35] James M. Hughes, Daniel J. Graham, and Daniel N. Rockmore. "Quantification of artistic style through sparse coding analysis in the drawings of Pieter Bruegel the Elder". In: *Proceedings of the National Academy of Sciences* 107 (2009), pp. 1279–1283. DOI: 10.1073/pnas.0910530107.

[36] H. W. Janson. *A History of Art*. Vol. 2. Harry N. Abrams, Inc., New York, 1978.

[37] C. Richard Johnson, Jr et al. "Image Processing for Artist Identification". In: *IEEE Signal Processing Magazine* 25.4 (2008), pp. 37–48. DOI: 0.1109/MSP.2008.923513.

[38] Yubin Kuang, David G. Stork, and Fredrik Kahl. "Improved curvature-based inpainting applied to fine art: Recovering van Gogh's partially hidden brush strokes". In: *Proceedings of SPIE*. Computer Vision and Image Analysis of Art II. Ed. by David G. Stork, Jim Coddington, and Anna Bentkowska-Kafel. Vol. 7869. 2011.

[39] Gitta Kutyniok, Jakob Lemvig, and Wang-Q Lim. "Shearlets and Optimally Sparse Approximations". In: *Shearlets. Multiscale Analysis for Multivariate Data*. Ed. by Gitta Kutyniok and Demetrio Labate. Applied and Numerical Harmonic Analysis. Springer, 2012, pp. 145–197. DOI: 10.1007/978-0-8176-8316-0_5. ISBN: 978-0-8176-8315-3.

[40] R. S. G. Lanzelotte et al. "The Portinari Project: Science and Art team up together to help cultural projects". In: *Proceedings of the 2nd International Conference on Hypermedia and Interactivity in Museums (ICHIM '93)*. The Museum Documentation Association. 1993, pp. 146–158.

[41] Jia Li and James Z. Wang. "Studying Digital Imagery of Ancient Paintings by Mixtures of Stochastic Models". In: *IEEE Transactions on Image Processing* 13 (2004), pp. 338–351. DOI: 10.1109/TIP.2003.821349.

[42] Jia Li, Lei Yao, Ella Hendriks, and James Z. Wang. "Rhythmic Brushstrokes Distinguish van Gogh from His Contemporaries: Findings via Automated Brushstroke Extraction". In: *IEEE Transactions on Pattern Analysis & Applications* (Oct. 2011). DOI: 10.1109/TPAMI.2011.203.

[43] Tony Lindeberg. "Feature Detection with Automatic Scale Selection". In: *International Journal of Computer Vision* 30.2 (1998), pp. 79–116. DOI: 10.1023/A:1008045108935.

[44] Morten Enemark Lund, Mark de Zee, and John Rasmussen. "Comparing calculated and measured curves in validation of musculuskeletal models". In: *XIII International Symposium on Computer Simulation in Biomechanics*. 2011.

[45] Siwei Lyu, Daniel Rockmore, and Hany Farid. "A digital technique for art authentication". In: *Proceedings of the National Academy of Sciences* 101.49 (Dec. 2004), pp. 17006–17010. DOI: 10.1073/pnas.0406398101.

[46] Laurens J. P. van der Maaten and Eric O. Postma. "Texton-based analysis of paintings". In: *Proceedings of SPIE*. Ed. by Andrew G. Tescher. Vol. 7798. Applications of Digital Image Processing XXXIII. 2010. DOI: 10.1117/12.863082.

[47] Guy P. Nason and Bernard W. Silverman. "The Stationary Wavelet Transform and some Statistical Applications". In: *Lecture Notes in Statistics*. Vol. 103. Springer-Verlag, 1995, pp. 281–300.

[48] Bruno A. Olshausen and Michael R. DeWeese. "Applied mathematics: The statistics of style". In: *Nature* 463 (Feb. 2010), pp. 1027–1028. DOI: 10.1038/4631027a.

[49] Bruno A. Olshausen and David J. Field. "Emergence of Simple-Cell Receptive Field Properties by Learning a Sparse Code for Natural Images". In: *Nature* 381 (June 1996), pp. 607–609. DOI: 10.1038/381607a0.

[50] Bruno A. Olshausen and David J. Field. "Sparse Coding with an Overcomplete Basis Set: A Strategy Employed by V1?" In: *Vision Research* 37.23 (Dec. 1997), pp. 3311–3325. DOI: 10.1016/S0042-6989(97)00169-7.

[51] Joe Palca. *Math Professor Helps Uncover Art Fakes*. Feb. 9, 2010. URL: http://www.npr.org/templates/story/story.php?storyId=123405424.

[52] Giuseppe Papari and Nicolai Petkov. "Adaptive Pseudo Dilation for Gestalt Edge Grouping and Contour Detection". In: *IEEE Transactions on Image Processing* 17.10 (Oct. 2008), pp. 1950–1962. DOI: 10.1109/TIP.2008.2002306.

[53] Hanchao Qi and Shannon Hughes. "A New Method for Visual Stylometry of Impressionist Paintings". In: *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. July 12, 2011, pp. 2036–2039. DOI: 10.1109/ICASSP.2011.5946912.

[54] Julie Rehmeyer. *Teaching a computer to spot a bogus Bruegel*. Jan. 11, 2010. URL: http://www.sciencenews.org/view/generic/id/53831/title/Math_Trek__Teaching_a_computer_to_spot_a_bogus_Bruegel.

[55] Camilla Stockmann. "Danske forskere afslører uægte kunstværker". In: *Politikken* (May 18, 2011). URL: http://politiken.dk/kultur/kunst/ECE1285493/danske-forskere-afsloerer-uaegte-kunstvaerker.

[56] David G. Stork. "Computer Vision and Computer Graphics Analysis of Paintings and Drawings: An Introduction to the Literature". In: *Computer Analysis of Images and Patterns*. Ed. by Xiaoyi Jiang and Nicolai Petkov. Vol. 5702. Lecture Notes in Computer Science. Springer, 2009, pp. 9–24. DOI: 10.1007/978-3-642-03767-2.

[57] David G. Stork and Alexander J. Kossolapov. "X-ray image analysis of Lorenzo Lotto's Husband and wife". In: *Proceedings of SPIE*. Computer Vision and Image Analysis of Art II. Ed. by David G. Stork, Jim Coddington, and Anna Bentkowska-Kafel. Vol. 7869. 2011. DOI: 10.1117/12.873191.

[58] Guilherme N. Teixeira, Raul Queiroz Feitosa, and Sidnei Paciornik. "Towards Automatic Painting Authentication". In: *Proceedings of the IAPR Conference on Machine Vision Applications*. 2002, pp. 530–533.

[59] *The Portinari Project*. URL: http://www.portinari.org.br.

[60] Martin Vetterli and Jelena Kovačević. *Wavelets and Subband Coding*. Prentice Hall, 1995. URL: http://www.waveletsandsubbandcoding.org.

[61] Martin Vetterli, Jelena Kovačević, and Vivek K. Goyal. *Fourier and Wavelet Signal Processing*. 2011. URL: http://www.fourierandwavelets.org.

[62] Qiong Wu et al. "Detection of digital doctoring in exemplar-based inpainted images". In: *International Conference on Machine Learning and Cybernetics*. Vol. 3. 2008, pp. 1222–1226. DOI: 10.1109/ICMLC.2008.4620591.

[63] Liang Ying and Ezzatollah Salari. "Beamlet Transform-Based Technique for Pavement Crack Detection and Classification". In: *Computer-Aided Civil and Infrastructure Engineering* 25 (8 2010), pp. 572–580. DOI: 10.1111/j.1467-8667.2010.00674.x.

[64] Mark de Zee et al. “Validation of a musculo-skeletal model of the mandible and its application to mandibular distraction osteogenesis”. In: *Journal of Biomechanics* 40.6 (2007), pp. 1192–1201. DOI: 10.1016/j.jbiomech.2006.06.024.

[65] Qiaoping Zhang and Isabelle Couloigner. “Accurate Centerline Detection and Line Width Estimation of Thick Lines Using the Radon Transform”. In: *IEEE Transactions on Image Processing* 16.2 (Feb. 2007), pp. 310–316. DOI: 10.1109/TIP.2006.88773.